

Copyright (c) 1993 - 2003 Compugen Ltd.

and is derived by analysis of the total score distribution.

OM nucleic - nucleic search, using sw model

Run on: April 9, 2003, 11:34:34 ; Search time 732 Seconds
 (without alignments)
 9546.674 Million cell updates/sec

Title: US-09-782-745-14

Perfect score: 538

Sequence: 1 ACGCCAGGGAGCTGTGAGGC.....CTGCAAAGAAAAAAA 538

Scoring table: Identity NUC Gapop 10-, Gapext 1.0

Searched: 1994485 seqs, 6494577260 residues

Total number of hits satisfying chosen parameters: 3988970

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl_NoHTG:
 1: gb_ba:*

2: gb_in:*

3: gb_om:*

4: gb_ov:*

5: gb_pat:*

6: gb_ph:*

7: gb_pi:*

8: gb_pr:*

9: gb_ro:*

10: gb_sts:*

11: gb_sy:*

12: gb_vn:*

13: gb_vt:*

14: em_ba:*

15: em_fun:*

16: em_num:*

17: em_in:*

18: em_mu:*

19: em_om:*

20: em_or:*

21: em_ov:*

22: em_dat:*

23: em_ph:*

24: em_pl:*

25: em_to:*

26: em_sts:*

27: em_un:*

28: em_vt:*

29: em_htg_hum:*

30: em_htg_inv:*

31: em_htg_other:*

32: em_htg_mus:*

33: em_htg_pdn:*

34: em_htg_sod:*

35: em_htg_mam:*

36: em_htg_vrt:*

37: em_sy:*

38: em_htgo_hum:*

39: em_htgo_mus:*

40: em_htgo_other:*

SUMMARIES

Result No. Score % Query Match length DB ID Description

1 538 100.0 538 5 AR028488 AR028488 Sequence 14

2 538 100.0 538 5 BC018052 BC018052 Sequence 14

3 533.2 99.1 551 8 AX285022 AX285022 Sequence

4 531.6 98.8 1245 5 HSU19143 HSU19143 Human GAGE-AF055473 Homo sapi

5 530 98.5 530 8 AF055473 Homo sapi

6 525.4 97.7 528 5 AR028490 AR028490 Sequence

7 512.8 95.3 540 5 AR028490 AR028490 Sequence

8 512.8 95.3 540 5 I55853 I55853 Sequence 16

9 512.2 95.2 637 8 BC024914 BC024914 Homo sapi

10 510.2 94.8 539 5 AR028492 AR028492 Sequence

11 510.2 94.8 539 5 I55855 I55855 Sequence 18

12 506 94.1 606 8 BC031628 BC031628 Homo sapi

13 504.8 93.8 532 5 AR028491 AR028491 Sequence

14 504.8 93.8 532 5 I55854 I55854 Sequence 17

15 501.4 93.2 528 5 AX334151 AX334151 Sequence

16 501.4 93.2 528 5 HSU19145 HSU19145 Human GAGE-AF058988 Homo sapi

17 498.8 92.7 527 8 HSU19147 HSU19147 Human GAGE-AF058988 Homo sapi

18 497.6 92.5 524 8 AF058988 AF058988

19 496.8 92.3 524 8 HSU19146 HSU19146 Human GAGE-AF055474 Homo sapi

20 498 90.7 526 8 AF055474 AF055474 Sequence

21 429.2 79.8 560 5 AR028489 AR028489 Sequence

22 429.2 79.8 560 5 I55852 I55852 Sequence 15

23 421.2 78.3 552 8 HSU19144 HSU19144 Human GAGE-AR028492 Sequence

24 374.4 69.6 646 5 AR028482 AR028482 Sequence

25 374.4 69.6 646 5 I55845 I55845 Sequence 1

26 374.4 69.6 646 5 HSU19142 HSU19142 Human GAGE-AF055474 Homo sapi

27 374.4 69.6 648 5 I57317 I57317 Sequence 1

c 28 301.4 56.0 530 5 AX284300 AX284300 Sequence

c 29 299.2 55.6 365 5 AX284692 AX284692 Sequence

c 30 209.6 39.0 662 8 BC004861 BC004861 Sequence

c 31 209.6 39.0 676 8 AF058989 AF058989

c 32 185.8 34.5 648 8 BC009232 BC009232 Homo sapi

c 33 183.8 34.2 642 5 AX359705 AX359705 Sequence

c 34 182.4 33.9 493 8 AJ318881 AJ318881 Homo sapi

c 35 179.8 33.4 620 8 HSA318880 HSA318880 Homo sapi

c 36 173.8 32.3 611 5 AX226501 AX226501 Sequence

c 37 163.2 30.3 79539 8 AC036664 AC036664 Homo sapi

c 38 156 29.0 580 5 AX078298 AX078298 Sequence

c 39 146.8 27.3 763 8 BC009230 BC009230 Homo sapi

c 40 140.8 26.2 20587 8 HSU185E6A HSU185E6A Human DNA s

c 41 140.8 26.2 62433 8 AL111391 AL111391 Human DNA

c 42 140.8 26.2 62493 8 AL096338 AL096338 Human DNA

c 43 131.4 24.4 951 8 AF055475 AF055475

c 44 130.4 24.2 475 5 AX226497 AX226497 Sequence

c 45 129.8 24.1 194418 8 AF235098 AF235098 Homo sapi

ALIGNMENTS

RESULT 1 AR028488 AR028488 Sequence 14 from patent US 5858689 .

LOCUS Sequence 14 ACCESSION AR028488 VERSION AR028488.1

KEYWORDS SOURCE Unknown. ORGANISM Unknown. UNCLASSIFIED Unknown.

REFERENCE 1 (bases 1 to 538)

AUTHORS van der Brugge, P., van den Bynde, B., Deinacker, O. and Boon-Falleur, T.

TITLE Isolated peptides derived from the gage tumor rejection antigen precursor and uses thereof

PATENT US 5858689-A 14 12-JAN-1999;

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

FEATURES source	Location/Qualifiers
BASE COUNT ORIGIN	1: .538 /organism="unknown" 160 a 116 c 155 g 107 t
Query Match Best Local Similarity Matches	100.0% ; Score 538; DB 5; Length 538; 100.0% ; Pred. No. 8.2e-124; Indels 0; Gaps 0, Conservative 0; Mismatches 0;
Qy	1 AGCCAGGGAGCTGAGGCACTGGTCTCGCTCCGACTTTCCCTC 60
Db	1 AGCCAGGGAGCTGAGGCACTGGTCTCGCTCCGACTTTCCCTC 60
Qy	61 ACTGAGATTAATCTGTGAAATACTGGTCTCGCTCCGACTTTCCCTC 120
Db	61 ACTGAGATTAATCTGTGAAATACTGGTCTCGCTCCGACTTTCCCTC 120
Qy	121 CAAAGCGTAACTGAGCTGAGATGGCTCTGAATGATTGGGCTATGGCTAGTG 180
Db	121 CAAAGCGTAACTGAGCTGAGATGGCTCTGAATGATTGGGCTATGGCTAGTG 180
Qy	121 CAAAGCGTAACTGAGCTGAGATGGCTCTGAATGATTGGGCTATGGCTAGTG 180
Db	121 CAAAGCGTAACTGAGCTGAGATGGCTCTGAATGATTGGGCTATGGCTAGTG 180
Qy	181 ATGAGTGGAAACCCAACCTGAGAAAGGGAAACAGCAAATCAAATCTGAGATCCCTG 240
Db	181 ATGAGTGGAAACCCAACCTGAGAAAGGGAAACAGCAAATCAAATCTGAGATCCCTG 240
Qy	241 CAGCTGCTAGGAGGAGGATGGGAGCTGAGCTGAGCTGAGCTAAC 120
Db	241 CAGCTGCTAGGAGGAGGATGGGAGCTGAGCTGAGCTGAGCTAAC 120
Qy	301 CTCTATGCCAGGAACGGTCACCCACAGCTGGTGTGAGTGTGAGATGGCTCTGATG 360
Db	301 CTCTATGCCAGGAACGGTCACCCACAGCTGGTGTGAGTGTGAGATGGCTCTGATG 360
Qy	361 GGAGGAGATGGCCCAAATCCAGGAGGGTGAAGAGGTGAAAG 420
Db	361 GGAGGAGATGGCCCAAATCCAGGAGGGTGAAGAGGTGAAAG 420
Qy	421 AATACAGCTTAAAGAAGACAATGGTCTATGTGAAATT 480
Db	421 AATACAGCTTAAAGAAGACAATGGTCTATGTGAAATT 480
Qy	481 TGTTCAATTAAAATCTCCAAATAAGCTTTAGCCCTCTGCAGAGGAAAG 538
Db	481 TGTTCAATTAAAATCTCCAAATAAGCTTTAGCCCTCTGCAGAGGAAAG 538
Qy	361 GGAGGAGATGGACCGCAAATAAGGAGGATGAGGAAACATGCTGAGCTGAGAAG 420
Db	361 GGAGGAGATGGACCGCAAATAAGGAGGATGAGGAAACATGCTGAGAAG 420
Qy	421 AATCACAGCTTAAAGAAGACAATGGTCTATGTGAAATT 480
Db	421 AATCACAGCTTAAAGAAGACAATGGTCTATGTGAAATT 480
Qy	481 TGTTCAATTAAAATCTCCAAATAAGCTTTAGCCCTCTGCAGAGGAAAG 538
Db	481 TGTTCAATTAAAATCTCCAAATAAGCTTTAGCCCTCTGCAGAGGAAAG 538
RESULT 2	
LOCUS	155851 Sequence 14 from patent US 5648226.
DEFINITION	Isolated peptides derived from tumor rejection antigens, and their use
ACCESSION	US 5648226-A 14-15-JUL-1997;
VERSION	1. .538 /organism="unknown"
KEYWORDS	Unknown.
SOURCE	Unknown.
ORGANISM	Unclassified.
REFERENCE	1 (bases 1 to 538) Van den Eynde, B., DeBacker, O. and Boon-Falleur, T. isolated peptides derived from tumor rejection antigens, and their use
AUTHORS	
TITLE	
JOURNAL	
FEATURES	
source	
BASE COUNT	160 a 116 c 155 g 107 t
ORIGIN	
Query Match Best Local Similarity Matches	100.0% ; Score 538; DB 5; Length 538; 100.0% ; Pred. No. 8.2e-124; Indels 0; Gaps 0, Conservative 0; Mismatches 0;
Qy	1 AGCCAGGGAGCTGAGGCACTGGTCTCGCTCCGACTTTCCCTC 60

Clone distribution: MGC clone distribution information can be found through the T.I.M.A.G.E. Consortium/LINN at: <http://image.llnl.gov>
 Series: IRAK Plate: 32 Row: k Column: 1 This clone was selected for full length sequencing because it

1 ACCCAGGGAGCTGAGGCACTGGTCTCGCTCCGACTTTCCCTC 60
 61 ACTGAGATCATCTGTGAAATACTGAGTGTGGAGAGATGAGCTGAGCTAGAC 120
 61 ACTGAGATCATCTGTGAAATACTGAGTGTGGAGAGATGAGCTGAGCTAGAC 120
 121 CAAAGCGTAACTGAGCTGAGATGGCTCTGAATGATTGGGCTATGGCTCTGAGT 180
 121 CAAAGCGTAACTGAGCTGAGATGGCTCTGAATGATTGGGCTATGGCTCTGAGT 180
 181 ATGAGTGGAAACCCAACCTGAGAAAGGGAAACAGCAAATCAAATCTGAGATCCCTG 240
 181 ATGAGTGGAAACCCAACCTGAGAAAGGGAAACAGCAAATCAAATCTGAGATCCCTG 240
 241 CAGCTGCTAGGAGGAGGATGGGAGCTGAGCTGAGCTGAGCTAAC 120
 241 CAGCTGCTAGGAGGAGGATGGGAGCTGAGCTGAGCTGAGCTAAC 120
 301 CTCTATGCCAGGAACGGTCACCCACAGCTGGTGTGAGTGTGAGATGGCTCTGATG 360
 301 CTCTATGCCAGGAACGGTCACCCACAGCTGGTGTGAGTGTGAGATGGCTCTGATG 360
 361 GGAGGAGATGGCCCAAATCCAGGAGGGTGAAGAGGTGAAAG 420
 361 GGAGGAGATGGCCCAAATCCAGGAGGGTGAAGAGGTGAAAG 420
 421 AATACAGCTTAAAGAAGACAATGGTCTATGTGAAATT 480
 421 AATACAGCTTAAAGAAGACAATGGTCTATGTGAAATT 480
 481 TGTTCAATTAAAATCTCCAAATAAGCTTTAGCCCTCTGCAGAGGAAAG 538
 481 TGTTCAATTAAAATCTCCAAATAAGCTTTAGCCCTCTGCAGAGGAAAG 538
 BC018052 BC018052 BC018052 BC018052 BC018052 BC018052
 LOCUS LOCUS LOCUS LOCUS LOCUS LOCUS
 DEFINITION Homo sapiens, G antigen 8, clone MGC: 26395 IMAGE: 4812462, mRNA, complete cds.
 ACCESSION BC018052 BC018052 BC018052 BC018052 BC018052 BC018052
 VERSION GI:17390105 GI:17390105 GI:17390105
 KEYWORDS MGC, MGC, MGC,
 SOURCE Homo sapiens. Homo sapiens. Homo sapiens.
 ORGANISM Bacteria; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Butheria; Primates; Catarrhini; Hominidae; Homo. NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 REFERENCE 1 (bases 1 to 551)
 AUTHORS Strausberg, R.
 TITLE Direct Submission
 JOURNAL Submitted (03-DEC-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 KEYWORDS NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 REMARK Contact: MGC help desk
 COMMENT Email: cgsbs-rc@mail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D
 Authors Preparation: Michael J. Brownstein (NHGRI) & Shiraki CDNA Library Preparation: Piero Carninci (RIKEN)
 Toshiyuki and
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNN)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www.shgc.stanford.edu>
 Contact: (Dickson, Mark) mcdf@paxil.stanford.edu
 Dickson, M., Schmitz, J., Grimmwood, J., Rodriguez, A., and Myers, R. M.

FEATURES source	passed the following selection criteria: matched mRNA gi: 4503878 . Location/Qualifiers 1 .. 551	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
/organism="Homo sapiens"		Lillie,J., Brown,J.L., Bolt,A., and van Huffel,C.
/db_xref="LocusID:26749"		Novel genes, compositions and methods for the identification, assessment, prevention, and therapy of human cancers
/clone="MGC:26395 IMAGE:4812462"		Patent: WO 017556 A 827 - 25-OCT-2001;
/tissue type="Brain, hippocampus"		Millennium Predictive Medicine, Inc. (US)
/clone Lib="NIH MGC_95"		Location/Qualifiers 1 .. 1245
/lab host="DH10B"		/organism="Homo sapiens"
/note="vector: pBluescript"		/db_xref="taxon:9606"
97. .447 /codon_start=1		BASE COUNT 382 a -308 c 390 g 153 t 12 others
/product="G antigen 8"		Query Match 98.9%; Score 531.6%; DB 5; Length 1245;
/protein_id="AAH18052..1"		Best Local Similarity 99.3%; Pred. No. 3..4..12..; Indels 0; Gaps 0;
/transref="GI:1739106"		FEATURES SOURCE
PATRQDPAAGDEASAGQPKPEADSSEQEGHPQTGCECSDGPQEMDPNPE		Qy 1 ACGCCAGGAGGTGTGAGCAGTGCTGGTCTGGTCGGACTTTTCCTCT 60
EVKTPPEGKEQSQC"		Db 179 ACGCAGGAGGTGTGAGCAGTGCTGGTCTGGTCGGACTTTTCCTCT 238
BASE COUNT 164 a 118 c 160 g 109 t		Query Match 98.9%; Score 531.6%; DB 5; Length 1245;
ORIGIN		Best Local Similarity 99.3%; Pred. No. 3..4..12..; Indels 0; Gaps 0;
Query Match 99.1%; Score 533.2; DB 8; Length 551;		FEATURES SOURCE
Best Local Similarity 99.4%; Pred. No. 1..3..e-12..; Indels 0; Gaps 0;		Qy 61 ACTGAGATTCACTGTGTGAAATAATGACTGTGGCAGGAGATCGACCTATGCCCTAGAC 120
Matches 535; Conservative 0; Mismatches 3; Indels 0; Gaps 0;		Db 239 ACTGAGATTCACTGTGTGAAATAATGACTGTGGCAGGAGATCGACCTATGCCCTAGAC 298
Qy 1 ACGCCAGGAGGTGTGAGCAGTGCTGGTCTGGTCGGACTTTTCCTCT 60		Qy 121 CAAGACGTACGTAGCCCTAGGCTTCCTGAATAATGATTGGGCTCTATGCCCTAGGCACTTCAGTG 180
Db 14 ACGCCAGGAGGTGTGAGCAGTGCTGGTCTGGTCGGACTTTTCCTCT 73		Db 299 CAAGACGTACGTAGCCCTCTGAATAATGATTGGGCTCTATGCCCTAGGCACTTCAGTG 358
Qy 61 ACTGAGATTCACTGTGTGAAATAATGACTGTGGCAGGAGATCGACCTATGCCCTAGAC 120		Qy 181 ATGAACTGTGAAACCGAACACCTGAGAAGGGAAACGAACTCAACGTGAGATCTGATCTG 240
Db 74 ACTGAGATTCACTGTGTGAAATAATGACTGTGGCAGGAGATCGACCTATGCCCTAGAC 133		Db 359 ATGAACTGTGAAACCGAACACCTGAGAAGGGAAACGAACTCAACGTGAGATCTGATCTG 418
Qy 121 CAAGACCTACTAGTAGGCCCTCCCTGAAATGATTGGCCCTATGCCCTGGCAGCTG 180		Qy 241 CAGGTGCTAGGAGGACAGGGTCACCCACAGGTGAGATGTGAAAGATGTGCTGAG 300
Db 134 CAAGACGTAACTAGACCTGTGAACTAGCCCTCCCTGAAATGATTGGCCCTATGCCCTGGCAGCTG 193		Db 419 CAGGTGCTAGGAGGACAGGGTCACCCACAGGTGAGATGTGAGATGTGCTGAG 478
Qy 181 ATGAACTGTGAAACCGAACACCTGAGAAGGGAAACCGAACACTCAACCTCAGATCCCTG 240		Qy 301 CTGATAGCCAGGACAGGGTCACCCACAGGTGAGATGTGAAAGATGTGCTGAG 360
Db 194 ATGAACTGTGAAACCGAACACCTGAGAAGGGAAACCGAACACTCAACCTCAGATCCCTG 253		Db 479 CTGATAGCCAGGACAGGGTCACCCACAGGTGAGATGTGAGATGTGCTGAG 538
Qy 241 CAGGTGCTAGGAGGAGGTGAGGAGCACTGTGAGTCAAGGGCCGAAGCTGAG 300		Qy 361 GGGAGGAGATGGACCAGCCGCAAATCCAGGAGGTGAAAGCTGAAAGCTGAAAGC 420
Db 254 CAGGTGCTAGGAGGAGGTGAGGAGCAATCCAGGAGGTGAGGAGCACTGTGAGTCAAGGGCCGAAGCTGAG 313		Db 539 GGGAGGAGATGGACCAGCCGCAAATCCAGGAGGTGAAAGCTGAAAGCTGAAAGC 598
Qy 301 CTCATAGGCCAAACAGGTCAACCCACAGACTGGTGTGAGTGTGGCTCTGATG 360		Qy 421 AATCACACTGTAAAGAGAACAGCTGGTGAATGATGAGGTGCTCTATGTGGAAATT 480
Db 314 CTGATAGCCAGAACAGGTCAACCCACAGACTGGTGTGAGTGTGGCTCTGATG 373		Db 599 AATCACACTGTAAAGAGAACAGGTCAACCCACAGACTGGTGTGAGTGTGGCTCTATGTGGAAATT 658
Qy 361 GGCAGGAGATGGACCAGCCAAATCCAGGAGGTGAGGAGCAAGCTGAAAGCTGAAAGC 420		Qy 481 TGTGATTAATAAATCTCCAAATAAGCTTACAGGCTTCTGAAAGAAAAAAA 538
Db 374 GGAGGAGATGGACCAGCCAAATCCAGGAGGTGAGGAGCAAGCTGAAAGCTGAAAGC 433		Db 659 TGTGATTAATAAATCTCCAAATAAGCTTACAGGCTTCTGAAAGAAAAAAA 716
Qy 421 AATCACACTGTAAAGAGAACAGCTGGTGAATGATGAGGAGCTGCTCCATGTGGAAATT 480		RESULT 5
Db 434 AATCACACTGTAAAGAGAACAGCTGGTGAATGATGAGGAGCTGCTCCATGTGGAAATT 493		HSU19143
Qy 481 TGTGATTAATCTCCAAATAAGCTTACAGGCTTCTGAAAGAAAAAAA 538		LOCUS HSU19143 530 bp mRNA
Db 494 TGTGATTAATCTCCAAATAAGCTTACAGGCTTCTGAAAGAAAAAAA 551		DEFINITION Human GAGE-2 protein mRNA, complete cds.
KEYWORDS human.		REFERENCE U19143 1 (bases 1 to 530)
SOURCE Homo sapiens		AUTHORS Van den Eynde,B., Peeters,O., De Backer,O., Gaugler,B., Lucas,S., and Boon,T.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		TITLE A new family of genes coding for an antigen recognized by autologous cytolytic T lymphocytes on a human melanoma
		J. Exp. Med. 182 (3), 689-698 (1995)
		JOURNAL 95378788
		MEDLINE

RESULT 11							ORGANISM	Homo sapiens
155855	155855	Sequence 18 from patent US 5648226.	539 bp	DNA	Linear	PAT 07-OCT-1997	REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 606)
LOCUS							AUTHORS	Strausberg, R.
DEFINITION							TITLE	Direct Submission
ACCESSION	155855						JOURNAL	(06-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A01, Bethesda, MD 20892-2590, USA
VERSION	155855.1	GI:2476649					REMARK	NIH-MGC Project URL: http://mgc.ncbi.nlm.nih.gov
KEYWORDS	Unknown,						COMMENT	Contact: MGCG help desk Email: cgaps-r@mail.nih.gov
SOURCE	Unknown,							Tissue Procurement: Life Technologies, Inc. cDNA Library Preparation: Life technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LILNL) DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
ORGANISM	Unclassified.							Center code: BCM-HGSC / Web Site: http://www.hgsc.bcm.tmc.edu/cDNA/ Contact: amg@bcm.tmc.edu
REFERENCE	1 (bases 1 to 539)	Van den Eynde, B., DeBacker, O., and Boon-Falleur, T.						GeneNarative, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M., Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Mizny, D.M., Richards, S., Gibbs, R.A.
AUTHORS		Isolated peptides derived from tumor rejection antigens, and their use						Clone distribution: MGCG clone distribution information can be found through the I.M.A.G.E. Consortium/INFINI at: http://image.llnl.gov Series: IRAK Place: 51 Row: b Column: 11 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4503876.
TITLE								Location/Qualifiers
FEATURES	Patent	US 5648226-A 18-15-JUL-1997;						
source		Location/Qualifiers						
BASE COUNT	158 a	113 c	157 g	111 t			Source	
ORIGIN								
Query Match	94.8%	Score 510.2; DB 5;	Length 539;					
Best Local Similarity	98.0%	Pred. No. 6.9e-117;						
Matches 528; Conservative 0; Mismatches 8; Indels 3; Gaps 1;								
Qy 3 GCGAGGGAGCTGCTGAGGCAAGCAGCTGCTGGACTCTTTTCCCTCTAC 62								
Db 1 GCGAGGGAGCTGCTGAGGCAAGCAGCTGCTGGACTCTTTTCCCTCTAC 60								
Qy 63 TGAGATTCACTGTGTGAATAATGAGTTGCCAGGAGATGCCACC---TATGGCTTAGA 119								
Db 61 TGAGATTCACTGTGTGAATAATGAGTTGCCAGGAGATGCCACCATTATGGCTTAGA 120								
Qy 120 CCAAGACGTACGTAGTACAGCTTCTGAAATATGTTGGCCATATCGGAGCTGAGTCTAGT 179								
Db 121 CCAAGGGCTPATGTACAGCTTCTGAAATCGGAGCTGAGTCTAGT 180								
Qy 180 GATGAGTGGACCAACACCTCTGAGAACGGCAACTACAGCTGGATTCT 239								
Db 181 GATGAGTGGACCAACACCTCTGAGAACGGCAACTACAGCTGGATTCT 240								
Qy 240 GCAGCTGTCTGGAGGGAGCTGGATCTGAGCTTCAAGGCCAAGCTCTGAA 299								
Db 241 GCAGCTGTCTGGAGGGAGCTGGATCTGAGCTTCAAGGCCAAGCTCTGAA 300								
Qy 300 GCTCATAGCCAGGAGGGTCACCCACAGACTGGGTCTGAGTGGTCTCTGAT 359								
Db 301 GCTGATAGCCGGACAGGTCACTCCACAGACTGGGTCTGAGTGGTCTCTGAT 360								
Qy 360 GGGCAGGAGTGGACCCGAATCCAGAGGGTGAAACCGCTGAAAGTGAAAG 419								
Db 361 GGGCAGGAGTGGACCCGAATCCAGAGGGTGAAACCGCTGAAAGTGAAAG 420								
Qy 420 CAATCACAGTGTAAAGAAACAGCTTGAATGATGTCAGCTTGTGAAAT 479								
Db 421 CAATCACAGTGTAAAGAAACAGCTTGAATGATGTCAGCTTGTGAAAT 480								
Qy 480 TTGTTCATTAATTCCTCCATTAATTAACCTTACAGCTTCTGCAAAGAAAAAAA 538								
Db 481 TTGTTCATTAATTCCTCCATTAATTAACCTTACAGCTTCTGCAAAGAAAAAAA 539								
RESULT 12								
LOCUS	BC031628							
DEFINITION	Homo sapiens, G antigen 7B, clone MGC:34597	606 bp	mRNA	linear	PRI 27-JUN-2002			
ACCESSION	BC031628							
VERSION	BC031628.1	GI:21619469						
KEYWORDS	MGC.							
SOURCE	human.							

LOCUS AX334151 528 bp DNA linear PAT 09-JAN-2002
 DEFINITION Sequence 4660 from Patent WO194629.
 ACCESSION AX334151
 VERSION GI:18124870
 KEYWORDS human.
 SOURCE ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1
 AUTHORS Young, P. E., Augustus, M., Carter, K. C., Ebner, R., Endress, G.,
 Horrigan, S., Soppet, D. R. and Weaver, Z.
 TITLE Cancer gene determination and therapeutic screening using signature
 Gene Sets
 JOURNAL Patent : WO 0194629-A 4660 13-DEC-2001;
 FEATURES Avalon Pharmaceuticals (US)
 Location/Qualifiers 1..528
 source /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 147 a 114 c 156 g 111 t
 ORIGIN

Query	Match	Score	DB	Length
	Best Local Similarity	93.2%	5	528;
	Matches	98.3%	Pred.	No. 1..114;
	518;	Conservative	0;	Mismatches
			6;	Indels
			3;	Gaps
Qy	2	CGCCAGGGAGCTGTGAGCAAGTGTCTGGCTGGCTCGGCTCGGACTCTTTCCTTA	61	
Db	1	CGCCAGGGAGCTGTGAGCAAGTGTCTGGCTGGCTGGCTCGGCTCGGACTCTTTCCTTA	60	
Qy	62	CTGAGATTCTCTGTGTGAATATGAGTTGGCAGGAAGATGCACC--TATCGGCCTAG	118	
Db	61	CTGAGATTCTCTGTGTGAATATGAGTTGGCAGGAAGATGCACC--TATCGGCCTAG	120	
Qy	119	ACCAAGACCTACGTAGACCTCTCTGAATGATGGCCTATGGGCCAGCATTGACCTTATTCGCTAG	178	
Db	121	ACCAAGGGCTATGTACGGCTCTCTGAATAATGATGGCCTATGGGCCAGCAGTTGAG	180	
Qy	179	TGATEAAGTGAAACCGAACACCTTGAGAAACGGAAACAGCAACTCAAGCTCAGATCC	238	
Db	181	TGATGAAAGTGAAACCGAACACCTTGAGAAACGGAAACAGCAACTCAAGCTCAGATCC	240	
Qy	239	TGCAGCTGCTCAGGGAGGAGGATGAGGAGGATCTGAGGTCAAAGGCCGAAGCTCTGA	298	
Db	241	TGCAGCTGCTCAGGGAGGAGGATCTGAGGTCAAAGGCCGAAGCTCTGA	300	
Qy	299	AGCTCATGGCAGGACAGGGTACCCACAGACTGGCTGAGCTGACAATGGCTCTGA	358	
Db	301	AGCTGATAGCAGGACAGGGTACCCACAGACTGGCTGAGCTGACAATGGCTCTGA	360	
Qy	359	TGGCAGGAGATGAACTGGCAATCCAGAGGGTGAAGAACGCTGAGGAAGGTGAAA	418	
Db	361	TGGCAGGAGATGAACTGGCAATCCAGAGGGTGAAGAACGCTGAGGAAGGTGAAA	420	
Qy	419	GCAATCACAGSTGTTAAAAGAACACGTTGAATGATGAGGCTGCTCTATGTTGAAA	478	
Db	421	GCAATCACAGSTGTTAAAAGAACACGTTGAATGATGAGGCTGCTCTATGTTGAAA	480	
Qy	479	TITGTTICATAAAATCTCCATAAAGCTTACAGCCTTCTGCATAAGTGGCTGCTCTATGTTGAAA	525	
Db	481	TITGTTICATAAAATCTCCATAAAGCTTACAGCCTTCTGCATAAGTGGCTGCTCTATGTTGAAA	527	



Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 9, 2003, 12:41:59 ; Search time 1040 Seconds
 (without alignments)
 8335.716 Million cell updates/sec

Title: US-09-782-745-14

Perfect score: 538
 Sequence: 1 AGCCAGGGAGCTGTGAGGC.....CTGCAAAAGAAAAAAA 538

Scoring table: IDENTITY_NUC
 Gapop 10.0 , Gapext 1.0

Searched: 60155 seqs, 8056825618 residues

Total number of hits satisfying chosen parameters: 120310

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : GenBank_HTG: * o
 1: gb_htg: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	DB ID	Description
c 1	163.2	30.3	240000	1	AC009528	Homo sapi
c 2	130.4	26.2	190309	1	AL645949	Homo sapi
c 3	131.4	24.4	48802	1	AC068431	Homo sapi
c 4	118	21.9	110000	1	AL831785	Homo sapi
c 5	97.2	18.1	48803	1	AC068431	Homo sapi
c 6	69.2	12.9	81953	1	AC016835	Homo sapi
c 7	69.2	12.9	93919	1	AC02415	Homo sapi
c 8	69.2	12.9	114344	1	AL445227	Homo sapi
c 9	69.2	12.9	114344	1	AL445227	Homo sapi
c 10	67.8	12.6	171489	1	AL772246	Homo sapi
c 11	67.6	12.6	180859	1	AC025553	Homo sapi
c 12	66	12.3	171489	1	AL772246	Homo sapi
c 13	64.6	12.0	180859	1	AC025553	Homo sapi
c 14	62.6	11.6	64891	1	AC118662	Homo sapi
c 15	60.6	11.3	93919	1	AC002415	Homo sapi
c 16	54.2	10.1	185032	1	AC109034	Rattus no
c 17	53.4	9.9	231755	1	AC115480	Rattus no
c 18	51.2	9.5	198806	1	AL671904	Mus muscu
c 19	50	9.3	33082	1	AC091365	Rattus no
c 20	49.6	9.2	188082	1	AC121784	Mus muscu
c 21	49.6	9.2	214690	1	AC083889	Mus muscu
c 22	49.6	9.2	221048	1	AC087329	Mus muscu
c 23	49.2	9.1	157392	1	AC120934	Rattus no
c 24	49.2	9.1	165210	1	AC101542	Mus muscu
c 25	49.2	9.1	196420	1	AC108842	Mus muscu
c 26	49	9.1	36077	1	AC106641	Rattus no
c 27	49	9.1	87750	1	AC111854	Rattus no
c 28	49	9.1	110141	1	AC114012	Rattus no
c 29	48.8	9.1	139655	1	AC125112	Mus muscu
c 30	48.6	9.0	110558	1	AC115492	Rattus no
c 31	48.2	9.0	63239	1	AC112945	Mus muscu
c 32	48	8.9	51267	1	AC100027	Mus muscu

ALIGNMENTS

RESULT 1	AC009528/C	AC009528	Homo sapiens chromosome 4, *** SEQUENCING IN PROGRESS ***, 21
LOCUS	AC009528	DEFINITION	unordered pieces.
ACCESSION	AC009528	VERSION	AC009528.7 GI:6604463
KEYWORDS	HTGS; HTGS_PHASEI.	ORGANISM	Homo sapiens
SOURCE	1 (bases 1 to 240000)	REFERENCE	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
ORGANISM	Direct Submission	AUTHORS	Mammalia; Butheria; Primates; Catarrhini; Hominidae; Homo.
COMMENT	Unpublished	TITLE	Submitted (26-AUG-1999) Department of Genetics, Stanford Human Genome Center, 855 California Avenue, Palo Alto, CA 94304, USA
NOTE:	On Dec 20, 1999 this sequence version replaced GI:573827.	REFERENCE	* NOTE: This is a working draft sequence. It currently
*	* consists of 21 contigs. The true order of the pieces	AUTHORS	* is not known and their order in this sequence record is
*	* arbitrary. Gaps between the contigs are represented as	TITLE	* runs of N, but the exact sizes of the gaps are unknown.
*	* This record will be updated with the finished sequence	JOURNAL	* as soon as it is available and the accession number will
*	* be preserved.	COMMENT	* 1416: contig of 1416 bp in length
*	*	*	1716: gap of unknown length
*	*	*	3005: contig of 1230 bp in length
*	*	*	3304: gap of unknown length
*	*	*	3305: contig of 1075 bp in length
*	*	*	4380: gap of unknown length
*	*	*	4678: gap of unknown length
*	*	*	6718: contig of 2040 bp in length
*	*	*	6719: gap of unknown length
*	*	*	7017: gap of unknown length
*	*	*	8520: contig of 1503 bp in length
*	*	*	8521: gap of unknown length
*	*	*	8820: 10657: contig of 1838 bp in length
*	*	*	10956: gap of unknown length
*	*	*	13220: contig of 2264 bp in length
*	*	*	13519: gap of unknown length
*	*	*	13221: contig of 1708 bp in length
*	*	*	15227: contig of 2274 bp in length
*	*	*	15526: gap of unknown length
*	*	*	15527: 18139: contig of 2613 bp in length
*	*	*	18140: 18438: gap of unknown length
*	*	*	18439: 20233: contig of 1795 bp in length
*	*	*	20234: 20532: gap of unknown length
*	*	*	20533: 22806: contig of 2274 bp in length
*	*	*	22807: 23104: gap of unknown length
*	*	*	23105: 24924: contig of 1820 bp in length
*	*	*	24925: 25222: gap of unknown length
*	*	*	25223: 27314: contig of 2092 bp in length

human.		
Homo sapiens		
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
*	27315	27612: gap of unknown length
*	27613	32471: contig of 4564 bp in length
*	32177	32474: gap of unknown length
*	32475	43341: contig of 10867 bp in length
*	43342	43639: gap of unknown length
*	43640	65292: contig of 21653 bp in length
*	65293	65590: gap of unknown length
*	65591	90191: contig of 24601 bp in length
*	90192	90489: gap of unknown length
*	90490	115831: contig of 25342 bp in length
*	115832	11619: gap of unknown length
*	116130	14584: contig of 27455 bp in length
*	143585	143882: gap of unknown length
*	143883	175459: contig of 31577 bp in length
*	175460	177575: gap of unknown length
*	1775758	240000: contig of 64243 bp in length.
1.	Location/Qualifiers	
1.	'240000	/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/chromosome="4"
BASE COUNT	61984 a 55073 c 54778 g 62188 t 5977 others	
ORIGIN		
Query Match	Score 163.2; DB 1; Length 240000;	
Best Local Similarity	70.0%; Pred. No. 1.3e-26;	
Matches	346; Conservative 0; Mismatches 113; Indels 35; Gaps 8;	
Qy	1 ACGCCAGGGCTGTGAGGCAGTGTGCTGCGACTCTTTTCCTCT 60	
Db	99869 AGAAAAGTGTCTGGGGACCTGTGAGGATATGCTCTCACATTCCGCGGTGACTCTT 99810	
Qy	61 ACTGAGATTCAATGTGTGAATAATGAGTTGGAGAAATGACCTATCGCTTAGAC 120	
Db	99809 ACTGAGACTCACTGTGTCA - ATGAGTCGTGAA - ATGAGTCGTGAA - ATGAGTCGTGCA - ATGAGTCGTGCA 99756	
Qy	121 CAAAGCCTACGTAGCTGAAACCTCTGAAATGTTGGCCCTATCGGCCGAGATGTTCTGCTATCTGCCAAGCTGTGTC 180	
Db	99755 CAATACAATCTGAACTAACCTAAAGG - - TGTCTGGCTTATCTGCCAAGCTGTGTC 99699	
Qy	161 ATGAGTCGTGAAACCTGAAAGGGAAACCTCAACGTCAGGATCCTG 240	
Db	99698 ATGAC - - - - - CACCTGAAAGAACCTGAACTCAAGCTGTGTC 99651	
Qy	241 CAGCTGTCTCAAGGGAGAGGATGAGGGACCATCTCGAGTCAGGCCAGCTGAAAG 300	
Db	99650 CACCTGCTCAGGGAGAGGAGCAAGCGGACATCTGCAGCTCAAAGCTGTGAAAG 99591	
Qy	301 CTCATAGCCTGGACACGGTCACCACAGCTGGTGTGACTGTGAGATGTCCTGATG 360	
Db	99590 CTGATTCGGACACTGGT - - - CAAAGACTGGTGTGACTGGATGATCTGATG 99535	
Qy	361 CGCAGGAGATGGACCCGGCAAAUTCCAGAGCAGGCTGAAAAGCCTGAAAGCTGAAAGC 420	
Db	99534 TGAAGGGGATGTGSCCTACTAAACCCAGAACCTGGAAATCTG - - CAAAGCTGAAAGC 99477	
Qy	421 ATCACAGTGTAAAGAGACAGTGTGAAATGATGAGGTGCTCTAATGTGAAAT 480	
Db	99476 ATTCAGGGT - - - - - TAAAGACATGCTGAAATGATGCGGGCTG - - CTATGTTGGGAT 99425	
Qy	481 TGTTCATTAAATT 494	
Db	99424 TGTTCATTCACACT 99411	
RESULT 2		
AL645949/c	AL645949 190309 bp DNA linear RP11-157G3, *** SEQUENCING IN	
LOCUS	Homo sapiens chromosome 4 clone RP11-157G3, *** in ordered pieces.	
DEFINITION	PROGRESS ***,	
ACCESSION	AL645949	
VERSION	AL645949.11 GI:22316144	
KEYWORDS	HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.	

Qy 299 AGCTCATGCCGAAAGCCGCTACCCACAGACTGGTGTGAGATGGTGTCCCTCA 358
 Db 102136 AGCTGATCTCGGGAGCTCTCAAGAC----TGGATGATGTGAGATGTCCTCA 102082
 Qy 359 TGGCAGGAGATGACCCGAAATTCCAGGAGGTGAAACGCTGAGAAGTGAAA 418
 Db 102081 T-GTCCGAGGAATTCTCCGARATCAGCAATTGAGGAGGTGAGG 102023
 Qy 419 GCATCACAGTGTAAAGAGAACGTTGAAATGATGAGGGTCTCTATGTTGGAAA 478
 Db 102022 GCAACCAAGGTTCAAGGAGAACATGGTAAACTGTTTATATTAGATA 101963
 Qy 479 TTGTTCTATTAAATTCTCCCAATAAGTTT 510
 Db 101962 CGTC--ACTTAAATAATCTCAAATCACGTTT 101934

RESULT 3
 AC068431 LOCUS AC068431 Homo sapiens chromosome 1 clone RP1-21H10 map 1, HTG 02-MAY-2000 DEFINITION LOW-PASS SEQUENCE SAMPLING.
 AC068431 VERSION AC068431.1 GI: 7677847
 KEYWORDS HTG; HTGS PHASE0.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens; Metazoa; Chordata; Craniata; Vertebrata; Butteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 48802)
 AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 TITLE Homo sapiens chromosome 1, clone RP11-21H10
 JOURNAL Unpublished 1 to 48802
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,V., Bastien,V., Beda,F., Boguslavsky,L., BoukigaBalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A., Dearalline,K., Dewar,K., Ditz,J.S., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Ferreira,P., FitzHugh,W., Gage,D., Dodge,S., Domino,M., Doyle,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Headford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lebocky,J., Levine,P., Lieu,C., Liu,G., Locke,P., Macdonald,I.P., Marquis,N., McCarthy,M., McEwen,P., McGurk,A., McKernan,K., McPhaeters,R., Meldrim,J., Menuez,L., Mihova,T., Mirzana,C., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Strange-Thomann,N., Stojanovic,N., Subramanian,A., Talama,J., Testafay,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,R., Vie,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

COMMENT All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE Direct Submission
 JOURNAL Submitted (02-MAY-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submission@genome.wi.mit.edu
 ----- Project Information
 Center project name: L7891
 Center clone name: 211_H_10

* NOTE: This record contains 59 individual sequencing reads that have not been assembled into

contigs. Runs of N are used to separate the reads and the order in which they appear is completely arbitrary. Low-pass sequence sampling is useful for identifying clones that may be gene-rich and allows overlap relationships among clones to be deduced. However, it should not be assumed that this clone will be sequenced to completion. In the event that the record is updated, the accession number will be preserved.
 * 1 715: contig of 715 bp in length
 * 716 815: gap of 100 bp
 * 816 1527: contig of 712 bp in length
 * 1528 1627: gap of 100 bp
 * 1628 2361: contig of 734 bp in length
 * 2362 2461: gap of 100 bp
 * 2462 3178: gap of 100 bp
 * 3179 3278: contig of 725 bp in length
 * 4004 4103: gap of 100 bp
 * 4104 4836: contig of 733 bp in length
 * 4837 4336: gap of 100 bp
 * 4937 5660: contig of 724 bp in length
 * 5661 5760: gap of 100 bp
 * 5761 6486: contig of 726 bp in length
 * 6487 6586: gap of 100 bp
 * 6587 7310: contig of 724 bp in length
 * 7311 7410: gap of 100 bp
 * 7411 8116: contig of 706 bp in length
 * 8117 8216: gap of 100 bp
 * 8217 8947: contig of 731 bp in length
 * 8948 9047: gap of 100 bp
 * 9048 9738: contig of 711 bp in length
 * 9739 9858: gap of 100 bp
 * 9859 10589: contig of 731 bp in length
 * 10590 10689: gap of 100 bp
 * 10690 11412: contig of 723 bp in length
 * 11413 11512: gap of 100 bp
 * 11513 12231: contig of 725 bp in length
 * 12238 12337: gap of 100 bp
 * 12338 13067: contig of 730 bp in length
 * 13068 13167: gap of 100 bp
 * 13168 13396: contig of 739 bp in length
 * 13397 14006: gap of 100 bp
 * 14007 14718: contig of 712 bp in length
 * 14719 14818: gap of 100 bp
 * 14819 15549: contig of 725 bp in length
 * 15544 15643: gap of 100 bp
 * 15644 16377: contig of 729 bp in length
 * 16373 16472: gap of 100 bp
 * 16473 172215: contig of 743 bp in length
 * 17216 17315: gap of 100 bp
 * 17316 18046: contig of 731 bp in length
 * 18047 18146: gap of 100 bp
 * 18147 18877: contig of 731 bp in length
 * 18878 18977: gap of 100 bp
 * 18978 19698: contig of 718 bp in length
 * 19696 19795: gap of 100 bp
 * 19795 20524: contig of 729 bp in length
 * 20525 20724: gap of 100 bp
 * 20625 21365: contig of 742 bp in length
 * 21367 21466: gap of 100 bp
 * 21467 22181: contig of 715 bp in length
 * 22182 22281: gap of 100 bp
 * 22282 22299: contig of 715 bp in length
 * 22297 23096: gap of 100 bp
 * 23097 23822: contig of 727 bp in length
 * 23824 23923: gap of 100 bp
 * 23924 24655: contig of 732 bp in length
 * 24656 24755: gap of 100 bp
 * 24755 25496: contig of 741 bp in length
 * 25497 25596: gap of 100 bp
 * 25597 26345: contig of 749 bp in length
 * 26346 26445: gap of 100 bp

27549 29549: contig of 2001 bp in length
 * 29550 29649: gap of 100 bp
 * 29550 32896: contig of 3247 bp in length
 * 32897 32996: gap of 100 bp
 * 32897 36277: contig of 3281 bp in length
 * 36278 36378: gap of 100 bp
 * 36378 38755: contig of 2378 bp in length
 * 38756 38855: gap of 100 bp
 * 38856 41582: contig of 2727 bp in length
 * 41583 41682: gap of 100 bp
 * 41683 44651: contig of 2969 bp in length
 * 44652 44751: gap of 100 bp
 * 44752 47834: contig of 3083 bp in length
 * 47835 47934: gap of 100 bp
 * 47935 59046: contig of 11112 bp in length
 * 59047 59146: gap of 100 bp
 * 59147 61468: contig of 2322 bp in length
 * 61469 61568: gap of 100 bp
 * 61569 63827: contig of 2259 bp in length
 * 63828 63928: gap of 100 bp
 * 63928 67325: contig of 3398 bp in length
 * 67326 67425: gap of 100 bp
 * 67426 69741: contig of 2316 bp in length
 * 69742 69841: gap of 100 bp
 * 69842 72033: contig of 2192 bp in length
 * 72034 72133: gap of 100 bp
 * 72134 74258: contig of 2125 bp in length
 * 74259 74358: gap of 100 bp
 * 74359 77178: contig of 2720 bp in length
 * 77079 77179: gap of 100 bp
 * 77179 80648: contig of 3470 bp in length
 * 80649 80748: gap of 100 bp
 * 80749 84081: contig of 3333 bp in length
 * 84082 84181: gap of 100 bp
 * 84182 87056: contig of 2875 bp in length
 * 87057 87156: gap of 100 bp
 * 87157 89197: contig of 2041 bp in length
 * 89198 89297: gap of 100 bp
 * 89298 91645: contig of 2348 bp in length
 * 91646 91745: gap of 100 bp
 * 91746 94371: contig of 2626 bp in length
 * 94372 94471: gap of 100 bp
 * 94472 96563: contig of 2092 bp in length
 * 96564 96663: gap of 100 bp
 * 96664 100313: contig of 3650 bp in length
 * 100314 100413: gap of 100 bp
 * 100414 102923: contig of 2510 bp in length
 * 102924 103023: gap of 100 bp
 * 103024 105754: contig of 2731 bp in length
 * 105755 105854: gap of 100 bp
 * 105855 108005: contig of 2151 bp in length
 * 115445 120013: contig of 4569 bp in length
 * 108006 108105: gap of 100 bp
 * 108106 113234: contig of 5129 bp in length
 * 113235 113344: gap of 100 bp
 * 113335 115344: contig of 2010 bp in length
 * 115345 115444: gap of 100 bp
 * 128008 130666: contig of 2659 bp in length
 * 130667 130766: gap of 100 bp
 * 120114 122698: contig of 2585 bp in length
 * 122699 122798: gap of 100 bp
 * 122799 122907: contig of 5109 bp in length
 * 127908 128007: gap of 100 bp
 * 128008 130666: contig of 2659 bp in length
 * 130667 130766: gap of 100 bp
 * 130767 1313234: contig of 2468 bp in length
 * 1313235 1333344: gap of 100 bp
 * 133335 136699: contig of 3365 bp in length
 * 136700 136799: gap of 100 bp
 * 136800 139489: contig of 2390 bp in length
 * 139490 139288: gap of 100 bp
 * 139290 142287: contig of 2998 bp in length
 * 142288 142388: gap of 100 bp
 * 142388 149385: contig of 6998 bp in length

* 149386 149485: gap of 100 bp
 * 149486 151937: contig of 2452 bp in length
 * 151938 152337: gap of 100 bp
 * 152038 154126: contig of 2089 bp in length
 * 154127 154226: gap of 100 bp
 * 154227 157753: contig of 3527 bp in length
 * 157754 157853: gap of 100 bp
 * 157854 159927: contig of 2074 bp in length
 * 159928 160027: gap of 100 bp
 * 160028 162267: contig of 2240 bp in length
 * 162268 162367: gap of 100 bp
 * 162368 167229: contig of 4862 bp in length
 * 167230 167329: gap of 100 bp
 * 167330 169663: contig of 2334 bp in length
 * 169664 169763: gap of 100 bp
 * 169764 172099: contig of 2336 bp in length
 * 172100 172199: gap of 100 bp
 * 172200 175654: contig of 3455 bp in length
 * 175655 175754: gap of 100 bp
 * 175755 180193: contig of 4439 bp in length
 * 180194 180393: gap of 100 bp
 * 180294 186167: contig of 5874 bp in length
 * 186168 186267: gap of 100 bp
 * 186268 188572: contig of 2305 bp in length
 * 188573 188672: gap of 100 bp
 * 188673 1914148: contig of 2746 bp in length
 * 191419 191518: gap of 100 bp
 * 191519 200088: contig of 8570 bp in length
 * 200089 200188: gap of 100 bp
 * 200189 20237: contig of 5049 bp in length
 * 20238 205317: gap of 100 bp
 * 205338 208097: contig of 2760 bp in length
 * 208098 208197: gap of 100 bp
 * 208198 211472: contig of 3275 bp in length
 * 211473 211572: gap of 100 bp
 * 211573 213169: contig of 5049 bp in length
 * 213370 213859: gap of 100 bp
 * 213870 216122: contig of 2253 bp in length
 * 216123 216222: gap of 100 bp
 * 216223 217243: contig of 2021 bp in length
 * 217244 219343: gap of 100 bp
 * 219344 2214613: contig of 2197 bp in length
 * 221370 224611: gap of 100 bp
 * 224611 224713: gap of 100 bp
 * 224714 224811: contig of 2098 bp in length
 * 226812 226911: gap of 100 bp
 * 226912 230682: contig of 3771 bp in length
 * 230683 230782: gap of 100 bp
 * 230783 233927: contig of 2445 bp in length
 * 233928 233027: gap of 100 bp
 * 233028 235051: contig of 2024 bp in length
 * 235052 235151: gap of 100 bp
 * 235152 237742: contig of 2491 bp in length
 * 237743 240388: contig of 2646 bp in length
 * 240389 240488: gap of 100 bp
 * 240489 242671: contig of 2183 bp in length
 * 242672 242771: gap of 100 bp
 * 242772 245338: contig of 2566 bp in length
 * 245338 245437: gap of 100 bp
 * 245438 248765: contig of 3328 bp in length
 * 248766 248865: gap of 100 bp
 * 248866 250924: contig of 2059 bp in length
 * 250925 251024: gap of 100 bp
 * 251025 253405: contig of 2381 bp in length

Query Match Score 118; DB 1; Length 110000;
 Best Local Similarity 96.0%; Pred. No. 1.4e-16;
 Matches 121; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 403 CTCAGAAGGTGAAAGCAATCAGTGTAAAGAGACACCTTGAAATGATGAGGGT 462
 Db 67467 CTGTTACAGGTGAAAGCAATCAGTGTAAAGAGCCGTGAAATGATGAGGGT 67526

QY 463 GCTCCATGGAAATTGTCATTAAATTCTCCCAATAAGCTTACAGCCUTCTGC 522
 Db 67527 GCTCCATGGAAATTGTCATTAAATTCTCCCAATAAGCTTACAGCCUTCTGC 67586

QY 523 AAAGGA 528
 Db 67587 AAAGGA 67592

RESULT 5
 LOCUS AC068431.1 48802 bp DNA linear HTG 02-MAY-2000
 DEFINITION Homo sapiens chromosome 1 clone RP1-211H10 map 1, LOW-PASS
 SEQUENCE SAMPLING
 AC068431.
 VERSION GI:76777847
 KEYWORDS HTG; HTGS; PHASE0.
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Bacteria; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 48802)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Homo sapiens chromosome 1, clone RP1-211H10
 Unpublished
 2 (bases 1 to 48802)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
 Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
 Campopiano,A., Castle,A., Chopeil,Y., Colangelo,M., Collins,S.,
 Collymore,A., Cooke,P., Dearallano,K., Dewar,K., Diaz,J.S.,
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W.,
 Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L.,
 Klein,J., Lakocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
 Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheters,R.,
 Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Milenga,V., Morrow,J.,
 Murphy,T., Naytor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 O'Neil,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,
 Pisani,C., Pollard,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
 Roy,A., Santos,R., Schaefer,S., Severy,P., Spencer,B.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
 Vassiliev,H., Vie,J., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE JOURNAL
 Submitted (02-MAY-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.P.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 Center Submission
 Center project name: L7691
 Center genome name: 211_H10

COMMENT
 NOTE: This record contains 59 individual
 * sequences that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

7946 8045: gap of 100 bp
 * 8046 8864: contig of 819 bp in length
 ** 8865 8964: gap of 100 bp in length
 *** 8965 9768: contig of 804 bp in length
 **** 9769 9868: gap of 100 bp
 ***** 9869 10879: contig of 811 bp in length
 ***** 10680 10779: gap of 100 bp
 ***** 10780 11155: contig of 766 bp in length
 ***** 11546 11645: gap of 100 bp
 ***** 11646 12460: contig of 815 bp in length
 ***** 12461 12560: gap of 100 bp
 ***** 12561 13372: contig of 812 bp in length
 ***** 13373 13472: gap of 100 bp
 ***** 13473 14215: contig of 803 bp in length
 ***** 14276 14375: gap of 100 bp
 ***** 14376 15173: contig of 798 bp in length
 ***** 15174 15273: gap of 100 bp
 ***** 15274 16048: contig of 785 bp in length
 ***** 16059 16158: gap of 100 bp
 ***** 16159 16553: contig of 795 bp in length
 ***** 16954 17053: gap of 100 bp
 ***** 17054 17840: contig of 787 bp in length
 ***** 17841 17940: gap of 100 bp
 ***** 17941 18750: contig of 810 bp in length
 ***** 18751 18850: gap of 100 bp
 ***** 18851 19656: contig of 806 bp in length
 ***** 19657 19756: gap of 100 bp
 ***** 19757 20551: contig of 805 bp in length
 ***** 20562 20661: gap of 100 bp
 ***** 21462 21446: contig of 795 bp in length
 ***** 21457 21556: gap of 100 bp
 ***** 21557 22482: contig of 826 bp in length
 ***** 22383 22482: gap of 100 bp
 ***** 22483 23306: contig of 824 bp in length
 ***** 23307 23406: gap of 100 bp
 ***** 23407 24217: contig of 811 bp in length
 ***** 24218 24317: gap of 100 bp
 ***** 24318 25116: contig of 799 bp in length
 ***** 25117 25216: gap of 100 bp
 ***** 25217 25399: contig of 783 bp in length
 ***** 26000 26099: gap of 100 bp
 ***** 26100 26899: contig of 780 bp in length
 ***** 26880 26979: gap of 100 bp
 ***** 26980 27773: contig of 794 bp in length
 ***** 27774 27873: gap of 100 bp
 ***** 27874 28668: contig of 795 bp in length
 ***** 28669 28768: gap of 100 bp
 ***** 28769 29552: contig of 824 bp in length
 ***** 29553 29692: gap of 100 bp
 ***** 29693 30523: contig of 831 bp in length
 ***** 30524 30623: gap of 100 bp
 ***** 30624 31423: contig of 800 bp in length
 ***** 31424 31523: gap of 100 bp
 ***** 31524 32321: contig of 798 bp in length
 ***** 32322 32421: gap of 100 bp
 ***** 32422 33244: contig of 823 bp in length
 ***** 33245 33344: gap of 100 bp
 ***** 33345 34172: contig of 828 bp in length
 ***** 34173 34272: gap of 100 bp
 ***** 34273 35019: contig of 807 bp in length
 ***** 35080 35179: gap of 100 bp
 ***** 35180 35967: contig of 788 bp in length
 ***** 35968 36067: gap of 100 bp
 ***** 36068 36846: contig of 779 bp in length
 ***** 36847 36946: gap of 100 bp
 ***** 36947 37740: contig of 794 bp in length
 ***** 37741 37841: gap of 100 bp
 ***** 38652 38751: gap of 100 bp
 ***** 38752 39566: contig of 815 bp in length
 ***** 39567 39666: gap of 100 bp
 ***** 40473 40472: contig of 806 bp in length
 ***** 40473 40572: gap of 100 bp

* 40573 41381: contig of 809 bp in length
 ** 41382 41481: gap of 100 bp
 *** 41482 42296: contig of 815 bp in length
 **** 42297 42396: gap of 100 bp
 ***** 42397 43213: contig of 817 bp in length
 ***** 43214 43313: gap of 100 bp
 ***** 43314 44124: contig of 811 bp in length
 ***** 44125 44224: gap of 100 bp
 ***** 44225 45038: contig of 804 bp in length
 ***** 45029 45128: gap of 100 bp
 ***** 45129 45931: contig of 809 bp in length
 ***** 45938 46037: gap of 100 bp
 ***** 46038 46814: contig of 777 bp in length
 ***** 46815 46914: gap of 100 bp
 ***** 46915 474710: contig of 796 bp in length
 ***** 47111 47810: gap of 100 bp
 ***** 47811 48580: contig of 770 bp in length
 ***** 48581 48680: gap of 100 bp
 ***** 48881 49465: contig of 785 bp in length
 ***** 49466 49565: gap of 100 bp
 ***** 49566 50377: contig of 812 bp in length
 ***** 50478 50477: gap of 100 bp
 ***** 50478 51260: contig of 783 bp in length
 ***** 51261 51360: gap of 100 bp
 ***** 51361 52143: contig of 783 bp in length
 ***** 52144 52243: gap of 100 bp
 ***** 52244 53049: contig of 806 bp in length
 ***** 53050 53149: gap of 100 bp
 ***** 53150 53951: contig of 810 bp in length
 ***** 53260 54059: gap of 100 bp
 ***** 54060 54833: contig of 773 bp in length
 ***** 54833 54932: gap of 100 bp
 ***** 54933 55712: contig of 780 bp in length
 ***** 55713 55812: gap of 100 bp
 ***** 55813 56660: contig of 792 bp in length
 ***** 56661 56704: gap of 100 bp
 ***** 56705 57487: contig of 783 bp in length
 ***** 57488 57587: gap of 100 bp
 ***** 57588 58395: contig of 808 bp in length
 ***** 58396 58495: gap of 100 bp
 ***** 58496 59302: contig of 807 bp in length
 ***** 59303 59402: gap of 100 bp
 ***** 59403 60205: contig of 803 bp in length
 ***** 60206 60305: gap of 100 bp
 ***** 60306 61119: contig of 814 bp in length
 ***** 61120 61219: gap of 100 bp
 ***** 61220 62026: contig of 807 bp in length
 ***** 62027 62126: gap of 100 bp
 ***** 62127 622941: contig of 815 bp in length
 ***** 622942 62041: gap of 100 bp
 ***** 63042 63857: contig of 816 bp in length
 ***** 63858 63957: gap of 100 bp
 ***** 63958 64772: contig of 815 bp in length
 ***** 64773 64772: gap of 100 bp
 ***** 64873 65675: contig of 803 bp in length

Query Match Score 69.2; DB 1; Length 81953;
 Best Local Similarity 70.8%; Pred. No. 9.9e-06;
 Matches 92; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

Qy 284 AGGGCGGAAGGCTGAGGTCAATGCCAGAACCTGGGTGTGATGTC 343
 Db 47098 AGTGCTTAACCTGGAGCTGATCTCAGGAGTCAGTCAAGCTGGATGATG 47157

Qy 344 TGAAGAGGT 413
 Db 47158 CGGAGTAGCTCTGATCTCCAGGGAGATTCTGCCAAANTCAGCHATTAAATGCC 47217

Qy 404 TGAAGAGGT 413
 Db 47218 AGAAGAGGT 47227

RESULT 7
 AC002415 AC002415 93919 bp DNA linear HTG 13-JUN-2002
 LOCUS Homo sapiens chromosome X clone bWXD142, *** SEQUENCING IN PROGRESS
 DEFINITION ***,
 **, 6 unordered pieces.
 ACCESSION AC002415
 VERSION AC002415_2 GI:21405635
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 93919)
 AUTHORS Chen, E., Brownstein,B.H., States,D.J., Schlessinger,D. and Mazzaella,R.
 TITLE Direct Submission
 JOURNAL Unpublished (1997)
 REFERENCE 2 (bases 1 to 93919)
 AUTHORS Brownstein,B.H., States,D.J. and Mazzaella,R.
 TITLE Direct Submission (12-AUG-1997) Center for Genetics in Medicine, Box 8232,
 Submitted (12-AUG-1997) Center for Genetics in Medicine, Box 8232,
 Washington University School of Medicine, 4566 Scott Avenue, St.
 Louis, MO 63110, USA
 On Jun 13, 2002 this sequence version replaced gi:2323256.
 Current status of this project is available at:
 'http://genome.wustl.edu/cgm/seq_projects.html',
 Submitted by:
 Ellison Chen,
 Advanced Center for Genetic Technology,
 Applied Biosystems Division of Perkin Elmer Corp.,
 850 Lincoln Center Drive,
 Foster City, CA 94404 USA
 e-mail: ellisonengenseq.apldbio.com
 and

Buddy Brownstein,
 Center for Genetics in Medicine,
 Washington University School of Medicine, Box 8232
 4566 Scott Avenue,
 St. Louis, MO 63110, USA
 e-mail: buddy@genetics.wustl.edu
 and

David J. States,
 Institute for Biomedical Computing
 Washington University in St. Louis
 700 South Euclid Ave.
 St. Louis, MO 63108 USA
 e-mail: states@ibc.wustl.edu.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 6 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 9560: contig of 9560 bp in length
 * 9561 9660: gap of 100 bp
 * 9661 25227: contig of 15567 bp in length
 * 25228 25327: gap of 100 bp
 * 25328 4575: contig of 20548 bp in length
 * 45876 45975: gap of 100 bp
 * 45976 73699: contig of 27724 bp in length
 * 73700 73799: gap of 100 bp
 * 73800 82475: contig of 8676 bp in length
 * 82476 82575: gap of 100 bp
 * 82576 93919: contig of 11344 bp in length.
 FEATURES Location/Qualifiers
 Source /organism="Homo sapiens"

/db_xref="taxon: 9606"
 /chromosome="X"
 /clones="bWXD142"
 BASE COUNT 25991 a 21189 c 20209 g 26030 t 500 others
 ORIGIN
 Query Match 12.9%; Score 69.2%; DB 1; Length 93919;
 Best Local Similarity 70.8%; Pred. No. 9.9e-06;
 Matches 92; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
 SOURCE
 Qy 284 AGGGCGGAAGCCCTGAAACTCATGCCAGAACAGTGGCTGTGACTG 343
 Db 59821 AGTCCTAACCTGGAACGTGATCTCCAGGTCTCACTCAAAGCTGGATGATG 59880
 Qy 344 TGAAGATGGTCTCTGATGGCAGAGATGACCCGCAATTCAAGAGGTAAAGCC 403
 Db 59881 CGGAGATAGTCCTGATGTCAGGGAAAGATTCTGCCAAATCAGGCAATTAAATGCC 59940
 Qy 404 TGAAGAGGT 413
 Db 59941 AGAAGGGGT 59950
 RESULT 8
 AL445227 AL445227 114344 bp DNA linear RP13-97115, *** SEQUENCING IN
 DEFINITION Homo sapiens chromosome X clone RP13-97115, *** SEQUENCING IN
 PROGRESS ***, 6 unordered pieces.
 ACCESSION AL445227
 VERSION AL445227.7 GI:11611365
 KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
 SOURCE human.
 ORGANISM Homo sapiens
 Bokaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 114344)
 AUTHORS McLay,K.
 TITLE Direct Submission
 JOURNAL Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 COMMENT On Dec 9, 2000 this sequence version replaced gi:11493333.
 Center: Sanger Centre
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquery@sanger.ac.uk
 Project Information
 Center project name: b99715
 Summary Statistics
 Assembly program: XGAP4; version 4.5
 Sequencing vector: Plasmid; L08752; 100% of reads
 Chemistry: Dye-Terminator Big Dye; 100% of reads
 Consensus quality: 11248 base at least Q40
 Consensus quality: 11345 bases at least Q30
 Consensus quality: 113597 bases at least Q20
 Insert size: 113844; sum-of-contigs
 Insert size: 122996; 9.8% error; agarose-fp
 Quality coverage: 5.9x in Q20 bases; sum-of-contigs Quality
 coverage: 5.72x in Q20 bases; agarose-fp
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 6 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 13809: contig of 13809 bp in length
 * 13810 13909: gap of 100 bp
 * 13910 51507: contig of 37538 bp in length
 * 51508 51607: gap of 100 bp

FEATURES

source

* 51608 65679: contig of 14072 bp in length
 * 65780 65779: gap of 100 bp
 * 65780 69061: contig of 3285 bp in length
 * 69065 69164: gap of 100 bp
 * 69165 101518: contig of 32354 bp in length
 * 101519 101618: gap of 100 bp
 * 101619 114344: contig of 12726 bp in length.

Location/Qualifiers

1. .114344 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="X"
 /clone="RP13-97115"
 /clone_id="RPCI-13.1"
 1. .13809 /note="assembly_fragment:00241
 fragment_chain:T
 clone_end:SP6
 vector_side:left"
 13910. .51507 /note="assembly_fragment:00680
 fragment_chain:T"
 51608. .65679 /note="assembly_fragment:00207
 fragment_chain:T"
 65780. .69064 /note="assembly_fragment:00300"
 69165. .101518 /note="assembly_fragment:01024"
 101619. .114344 /note="assembly_fragment:01825
 clone_end:T7
 vector_side:right"
 BASE COUNT 32658 a 24338 c 24212 g 32616 t 500 others
 ORIGIN

Query Match 12.9% Score 69.2; DB:1; Length 114344;
 Best Local Similarity 70.8%; Pred. No. 9.9e-06;
 Matches 92; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

Qy 284 AGGGCGGAAGCTGAAAGCTCATAGCCAGAACAGGTCACCACAGACTGGGTGAGTGAGT 343
 Db 39620 AGTCCTCTACCTGGAGCTATCCTGGAGCTATCCTGGAGCTGTCTGTAAGACTGGGTGAGT 39679

Qy 344 TGAGATGCTCTGATGGAGGATGGAGGATGCCAAATCCAGAGGAGGTGAAACGCC 403
 Db 39680 CGGAGATACTCTGATGTCCAGGGNAGATCTGCCAAAATCAGAGCHATTAAATGCC 39739

Qy 404 TGAAGAAGGT 413
 Db 39740 AGAGGGAGGT 39749

RESULT 9
 LOCUS AL445227 114344 bp DNA linear HTG 10-JUL-2001
 DEFINITION Homo sapiens chromosome X clone RP13-97115, *** SEQUENCING IN PROGRESS *** , 6 unordered pieces.
 ACCESSION AL445227
 VERSION AL445227.7 GI:11611395
 KEYWORDS HTGS; HTGS_PHASE1; HTGS_CANCELLED.
 SOURCE human
 ORGANISM Homo sapiens
 Mammalia; Butheria; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Butheria; Primates; Catarrhini; Hominidae; Homo.
 1. (bases 1 to 114344)

REFERENCE McClay K.
 AUTHORS McClay K.
 TITLE Direct Submission
 JOURNAL Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire, CB1 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerquests@sanger.ac.uk
 On Dec 9, 2000 this sequence version replaced gi:11493333.
 COMMENT Genome Center

BASE COUNT 32658 a 24338 c 24212 g 32616 t 500 others
 ORIGIN

Query Match 12.9%; Score 69.2; DB:1; Length 114344;
 Best Local Similarity 70.8%; Pred. No. 9.9e-06;
 Matches 92; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

Qy 284 AGGGCGGAAGCTGAAAGCTCATAGCCAGAACAGGTCACCACAGACTGGGTGAGTGAGT 343
 Db 92320 AGTCCTCTACCTGGAGCTATCCTGGAGCTGTCTGTAAGACTGGGTGAGT 92261

Qy 344 TGTAGATGCTCCTGATGGCAGGAGATGCCGCCAAATCCAGGGTGANACCC 403
 Db 92220 TGGAATATGTCCTGATGCCGGAGATTCTGCCAAATCAAGAACATTAAATSCC 92201

Qy 404 TGTAGAAGGT 413
 Db 92220 AGAGGAAGT 92191

RESULT 1.0
 LOCUS AL772246 171489 bp DNA linear HTG_13-AUG-2002
 DEFINITION Homo sapiens chromosome X clone RP13-34L8, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
 ACCESSION AL772246
 VERSION AL772246.4 GI:22265477
 KEYWORDS HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_FULLTOP;
 SOURCE human.
 ORGANISM Homo sapiens
 Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 180859)
 AUTHORS Waterston, R.H.
 TITLE The sequence of Homo sapiens clone
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 180859)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (10-MAR-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
 On Sep 1, 2000 this sequence version replaced gi:9954853.

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc/index.shtml
 Center project name: H_NH045B17
 ----- Summary Statistics -----
 Sequencing vector: M13; 100%
 Sequencing vector: Plasmid; 0%
 Chemistry: Dye-primer ET; 100% of reads
 Assembly program: Phrap; version 0.990119
 Consensus Quality: 170539 bases at least Q40
 Consensus Quality: 173308 bases at least Q30
 Consensus Quality: 174149 bases at least Q20
 Insert size: 178075; sum-of-contigs
 Insert size: 186799; 2.2% error; agarose-fp
 Quality Coverage: 12.50x in Q20 bases; agarose-fp
 coverage: 11.91x in Q20 bases; agarose-fp
 coverage: 5.22 in Q20 bases; sum-of-contigs
 Quality Coverage: 4.89 in Q20 bases; agarose-fp
 coverage: 4.89 in Q20 bases; sum-of-contigs
 Quality Coverage: 5.22 in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently consists of 23 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved. -----

* NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced.
 * The finished sequence as soon as it is available and by the accession number will be preserved.

FEATURES

source
 /organism="Homo sapiens"
 /ab_xref="taxon:9606"
 /chromosome="X"
 /clone="RP13-34L8"
 /clone.lib="RPCI-13.1"
 misc_feature
 1..171489
 /note="assembly fragment:02999"
 BASE COUNT 48292 a 37754 c 37294 g 48149 t
 ORIGIN

Query Match 12.6%; Score 67.8; DB 1; Length 171489;
 Best Local Similarity 6.2%; Pred. No. 2e-05; Indels 0; Gaps 0;
 Matches 75; Conservative 0; Mismatches 12;

Qy 76 TGTGAAATAGTGTGGCGGAGAATGCCCTATGCCCTAGACAGCGCTACGTAG 135
 Db 95798 TGTGAAATAGTGAATGGCTAGCAAGATCACATAGGCTAGCAAGATGATGTAC 95739
 Qy 136 AGCTCTGAAATGATGGCCTATGC 162

FEATURES

source

1. * 12396 13579: contig of 1184 bp in length
 * 13580 13679: gap of unknown length
 * 13680 15201: contig of 1522 bp in length
 * 15202 15301: gap of unknown length
 * 15302 16390: contig of 1089 bp in length
 * 16391 16490: gap of unknown length
 * 16491 18667: contig of 2177 bp in length
 * 18668 18767: gap of unknown length
 * 18768 20870: contig of 2102 bp in length
 * 20870 20669: gap of unknown length
 * 23568: contig of 2599 bp in length
 * 23569 23668: gap of unknown length
 * 23669 24252: contig of 584 bp in length
 * 24253 24252: gap of unknown length
 * 24353 28943: contig of 4591 bp in length
 * 28944 29044: gap of unknown length
 * 37619: contig of 8576 bp in length
 * 37620 37719: gap of unknown length
 * 37720 49220: contig of 12208 bp in length
 * 49228 50227: gap of unknown length
 * 50228 69100: contig of 19073 bp in length
 * 69101 69201: gap of unknown length
 * 69201 90261: contig of 21060 bp in length
 * 90261 90361: gap of unknown length
 * 90361 141186: contig of 51456 bp in length
 * 141186 141187: gap of unknown length
 * 141187 180819: contig of 38973 bp in length.
 . Location/Qualifiers

1. * 180819

/organism="Homo sapiens"
/db_xref="taxon: 9606"
/chromosome="X"
/clone="RP11-485B17"

misc_feature

1. * 1341
/note="assembly_name:Contig19"
1442. * 2824
/note="assembly_name:Contig24"
2925. * 4091
/note="assembly_name:Contig25"
4192. * 5300
/note="assembly_name:Contig26"
5401. * 6531
/note="assembly_name:Contig40"
6612. * 1946
/note="assembly_name:Contig44"
8047. * 9816
/note="assembly_name:Contig45"
9917. * 11066
/note="assembly_name:Contig46"
11167. * 12295
/note="assembly_name:Contig47"
12396. * 13579
/note="assembly_name:Contig48"
13680. * 15201
/note="assembly_name:Contig49"
15302. * 16390
/note="assembly_name:Contig50"
16491. * 18667
/note="assembly_name:Contig52"
18768. * 20869
/note="assembly_name:Contig53"
20970. * 23568
/note="assembly_name:Contig54"
clone_end:SP6
vector_side:right"
23569. * 24252
/note="assembly_name:Contig6"
24333. * 28943
/note="assembly_name:Contig55"
29044. * 37619
/note="assembly_name:Contig56"
37720. * 49927
/note="assembly_name:Contig57"

misc_feature

50028. * 69100
/note="assembly_name:Contig58"
69201. * 90260
/note="assembly_name:Contig59"
90361. * 141786
/note="assembly_name:Contig60"
141887. * 180819
/note="assembly_name:Contig61"
53372 a 37170 c 37454 g 50639 t 2224 others
BASE COUNT
ORIGIN

Query Match Score 67.6%;
Best Local Similarity 70.0%;
Matches 91; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy 284 AGGCAGCGAAGCCTAGGAGCTAGGAGGTACACCAAGACTGGGGTGTGAGTG 343
Db 93228 AGTCGCTGAGCTGAGGTATCTCCAGGAGCTGTCAGTCAGCTAAAGCTGGGGATGATG 93287

Qy 344 TGAATGTTGCTGATGGGGAGATGGACGCCAAATCCAGGAGGTAAACGCC 403
Db 93288 CGGATGGTCCTGAGTCCAGGGAGAGATCTGACAAAGTGTAGCAATTAAATGCC 93347

Qy 404 TGAGAAGAGT 413
Db 93348 AGAAGAGGT 93357

RESULT 12
AL772246 LOCUS AL772246
DEFINITION Homo sapiens chromosome X clone RP13-34L8, *** SEQUENCING IN PROGRESS ***, in ordered pieces.

ACCESSION AL772246
VERSION AL772246.4
KEYWORDS GI:22265477 HG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 171489)
AUTHORS Whitehead, S
TITLE Direct Submission
JOURNAL Submitted (10-AUG-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail: enquires@sanger.ac.uk
COMMENT On Aug 15, 2002 this sequence version replaced gi:22002760.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bB4L8
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 171482 bases at least Q40
Consensus quality: 171488 bases at least Q30
Consensus quality: 171489 bases at least Q20
Insert size: 171489; sum-of-contigs
Insert size: 186799; 2.2% error; agarose-fp
Quality coverage: 12.50x in Q20 bases; sum-of-contains Quality coverage: 11.1x in Q20 bases; agarose-fp
coverage: 11.1x in Q20 bases;

* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES source
1. * 171489
/organism="Homo sapiens"
/db_xref="taxon: 9606"

/chromosome="X"
 /clone="RP13-341L8"
 /clone_id="RPCI-3.1"
 misc_feature
 1. .171489
 /note="assembly fragment:02999"
 BASE COUNT 48292 a 37754 c 37294 g 48149 t
 ORIGIN
 Query Match 12.3%; Score 66; DB 1; Length 171489;
 Best Local Similarity 69.2%; Pred. No. 5e-05; Indels 0; Gaps 0;
 Matches 90; Conservative 0; Mismatches 40; Gap 0;
 Qy 284 AGGGCGAAGCCCTAAAGCTATAGCCAGGACAGGGTCAACCACAGACTGGGTGAGTG 343
 Db 81057 AGTGCCTGACCTGCTGAGCTATCAGGAGCTGTCATCAAGACTGGGTGAGTG 81116
 Qy 344 TGAAGATGCCATGGGAGGAGATGGACCCGCAATTCAAGAGGTGGTGAACGCC 403
 Db 81117 CGGAGATGTCCTGATATCAGGGAAAGATCTGCCAAATACTAGGCAATTAAATGCC 81176
 Qy 404 TGAGAAGGT 413
 Db 81177 AGAGGAGGT 81186

RESULT 13
 AC025553/c
 LOCUS AC025553 180859 bp DNA linear HTG 01-SEP-2000
 DEFINITION HOMO SAPIENS chromosome X clone RP11-485B17, WORKING DRAFT
 SEQUENCE, 23 unordered pieces.
 AC025553
 AC025553.5 GI:958270
 VERSION HTGS_PHASE1; HTGS_DRAFT.
 KEYWORDS SOURCE
 ORGANISM Homo sapiens.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarhini; Hominidae; Homo.
 1 (bases 1 to 180859)
 AUTHORS Waterston, R.H.
 TITLE The sequence of Homo sapiens clone
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 180859)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (10-MAR-2000) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 On Sep 1, 2000 this sequence version replaced gi:9954853.
 COMMENT -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site:<http://Genome.wustl.edu/gsc/index.shtml>
 ----- Project Information -----
 Center project name: H_NHG485B17
 ----- Summary Statistics -----
 Sequencing vector: M13; 100%
 Sequencing vector: plasmid; 0%
 Chemistry: Dye-Primer ET; 100% of reads
 Chemistry: Dye-terminator Big Dye; 0% of reads
 Assembly program: Phrap; Version 0.900319
 Consensus quality: 170599 bases at least Q40
 Consensus quality: 173308 bases at least Q30
 Consensus quality: 174949 bases at least Q20
 Insert size: 18000; agarose-fp
 Insert size: 170705; sum-of-contigs
 Quality coverage: 4.89 in Q20 bases; agarose-fp
 Quality coverage: 5.22 in Q20 bases; sum-of-contigs

FEATURES
 source
 1. .180859
 /organism="Homo sapiens"
 /db_xref="taxon: 9606"
 /chromosome="X"
 /clone="RP11-485B17"

misc_feature
 1. .1341
 /note="assembly_name:Contig19"
 /note="assembly_name:Contig24"
 /note="assembly_name:Contig25"
 /note="assembly_name:Contig26"
 /note="assembly_name:Contig40"
 /note="assembly_name:Contig46"
 /note="assembly_name:Contig44"
 /note="assembly_name:Contig45"
 /note="assembly_name:Contig47"

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 23 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1. .1341 contig of 1341 bp in length
 1. .1342 gap of unknown length
 1. .1442 contig of 1383 bp in length
 1. .2825 gap of unknown length
 1. .2925 contig of 1167 bp in length
 1. .4091 gap of unknown length
 1. .4092 gap of unknown length
 1. .4192 contig of 1109 bp in length
 1. .5300 gap of unknown length
 1. .5401 contig of 1131 bp in length
 1. .6531 gap of unknown length
 1. .6532 contig of 1315 bp in length
 1. .7946 gap of unknown length
 1. .7947 gap of unknown length
 1. .8046 gap of unknown length
 1. .9816 contig of 1770 bp in length
 1. .9917 gap of unknown length
 1. .11066 contig of 1150 bp in length
 1. .11067 gap of unknown length
 1. .11295 contig of 1129 bp in length
 1. .11167 gap of unknown length
 1. .12295 gap of unknown length
 1. .12396 contig of 1184 bp in length
 1. .13579 gap of unknown length
 1. .13679 gap of unknown length
 1. .13680 contig of 1522 bp in length
 1. .15201 gap of unknown length
 1. .15301 gap of unknown length
 1. .15302 contig of 1089 bp in length
 1. .16390 gap of unknown length
 1. .16490 gap of unknown length
 1. .18667 contig of 2177 bp in length
 1. .18668 gap of unknown length
 1. .18768 contig of 2102 bp in length
 1. .20969 gap of unknown length
 1. .20870 gap of unknown length
 1. .23568 contig of 2599 bp in length
 1. .23569 gap of unknown length
 1. .23668 gap of unknown length
 1. .24252 contig of 584 bp in length
 1. .24253 gap of unknown length
 1. .24352 contig of 4591 bp in length
 1. .28943 contig of 19073 bp in length
 1. .29043 gap of unknown length
 1. .29044 gap of unknown length
 1. .37619 gap of unknown length
 1. .37720 gap of 12208 bp in length
 1. .49927 gap of unknown length
 1. .50027 gap of unknown length
 1. .69100 gap of unknown length
 1. .69200 gap of unknown length
 1. .69201 contig of 21060 bp in length
 1. .90261 gap of unknown length
 1. .90361 contig of 51426 bp in length
 1. .141787 gap of unknown length
 1. .141886 gap of unknown length
 1. .180859 contig of 38973 bp in length.

13988 14087: gap of 100 bp
 * 14088 14769: contig of 682 bp in length
 ** 14869 15548: gap of 100 bp
 *** 14870 15548: contig of 679 bp in length
 **** 15549 15648: gap of 100 bp
 ***** 15649 16224: contig of 676 bp in length
 ***** 16325 16424: gap of 100 bp
 ***** 16425 17108: contig of 684 bp in length
 ***** 17109 17208: gap of 100 bp
 ***** 17209 17392: contig of 684 bp in length
 ***** 17893 17992: gap of 100 bp
 ***** 17993 18672: contig of 680 bp in length
 ***** 18673 18772: gap of 100 bp
 ***** 18773 1975: contig of 703 bp in length
 ***** 19476 19575: gap of 100 bp
 ***** 19576 20210: contig of 695 bp in length
 ***** 20271 20370: gap of 100 bp
 ***** 20371 21027: contig of 657 bp in length
 ***** 21028 21127: gap of 100 bp
 ***** 21128 21804: contig of 677 bp in length
 ***** 21805 21904: gap of 100 bp
 ***** 21905 22597: contig of 693 bp in length
 ***** 22598 22697: gap of 100 bp
 ***** 22698 23391: contig of 694 bp in length
 ***** 23392 23491: gap of 100 bp
 ***** 23492 24200: contig of 709 bp in length
 ***** 24201 24300: gap of 100 bp
 ***** 24301 24983: contig of 683 bp in length
 ***** 24984 25083: gap of 100 bp
 ***** 25084 25777: contig of 694 bp in length
 ***** 25778 25877: gap of 100 bp
 ***** 25878 26566: contig of 683 bp in length
 ***** 26561 26660: gap of 100 bp
 ***** 26561 27335: contig of 699 bp in length
 ***** 27360 27459: gap of 100 bp
 ***** 28128 28128: contig of 669 bp in length
 ***** 28129 28228: gap of 100 bp
 ***** 28229 28896: contig of 668 bp in length
 ***** 28897 28996: gap of 100 bp
 ***** 28997 29663: contig of 667 bp in length
 ***** 29664 29763: gap of 100 bp
 ***** 29764 30454: contig of 691 bp in length
 ***** 30455 30554: gap of 100 bp
 ***** 31232 31232: contig of 678 bp in length
 ***** 31233 31332: gap of 100 bp
 ***** 32029 32128: contig of 696 bp in length
 ***** 32129 32816: gap of 100 bp
 ***** 32817 32916: gap of 100 bp
 ***** 32917 33595: contig of 679 bp in length
 ***** 33596 33655: gap of 100 bp
 ***** 33695 34380: contig of 685 bp in length
 ***** 34381 34480: gap of 100 bp
 ***** 34481 35189: contig of 709 bp in length
 ***** 35190 35259: gap of 100 bp
 ***** 35290 35957: contig of 668 bp in length
 ***** 35958 36057: gap of 100 bp
 ***** 36058 36734: contig of 677 bp in length
 ***** 36735 36834: gap of 100 bp
 ***** 36835 37502: contig of 668 bp in length
 ***** 37503 37603: gap of 100 bp
 ***** 37603 38286: contig of 684 bp in length
 ***** 38287 38386: gap of 100 bp
 ***** 38387 39060: contig of 674 bp in length
 ***** 39061 39161: gap of 100 bp
 ***** 39161 39850: contig of 690 bp in length
 ***** 39851 39951: gap of 100 bp
 ***** 40657 40755: gap of 100 bp
 ***** 40757 41452: contig of 696 bp in length
 ***** 41453 41552: gap of 100 bp
 ***** 41553 42265: contig of 708 bp in length
 ***** 42261 42360: gap of 100 bp

42361 43028: contig of 668 bp in length
 * 43029 43128: gap of 100 bp
 ** 43129 43800: contig of 672 bp in length
 *** 43801 43900: gap of 100 bp
 **** 44566: contig of 666 bp in length
 ***** 44567 44666: gap of 100 bp
 ***** 44667 45354: contig of 688 bp in length
 ***** 45355 45154: gap of 100 bp
 ***** 45154 46131: contig of 683 bp in length
 ***** 45555 46237: gap of 100 bp
 ***** 46238 46917: contig of 680 bp in length
 ***** 46918 47017: gap of 100 bp
 ***** 47018 47676: contig of 659 bp in length
 ***** 47677 47776: gap of 100 bp
 ***** 47777 48461: contig of 685 bp in length
 ***** 48462 48561: gap of 100 bp
 ***** 48562 49248: contig of 687 bp in length
 ***** 49249 49348: gap of 100 bp
 ***** 49349 50024: contig of 676 bp in length
 ***** 50025 50124: gap of 100 bp
 ***** 50125 50802: contig of 678 bp in length
 ***** 50803 50902: gap of 100 bp
 ***** 50903 51607: contig of 705 bp in length
 ***** 51608 51707: gap of 100 bp
 ***** 51708 52385: contig of 678 bp in length
 ***** 52386 52485: gap of 100 bp
 ***** 52486 53131: contig of 646 bp in length
 ***** 53132 53201: gap of 100 bp
 ***** 53232 53905: contig of 674 bp in length

Query Match Score 62.6; DB 1; Length 64891;
 Best Local Similarity 66.9%; Pred. No. 0.00029;
 Matches 89; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Qy 281 TCAAGGGCCGAAAGCTGAAGGTCTATGCCAGGTGGGTCAACAGTGGTGTGA 340
 Locus Homo sapiens chromosome X clone bWXD142, *** SEQUENCING IN PROGRESS
 Definition AC002415
 * *, 6 unordered pieces.
 Db 32401 TTAAGTGTCTGACTCTGGAGTCATCTCAGGATTGTCTAGTCAGTCAAGTCTGGATGATGA 32342
 Accession AC002415
 Version AC002415.2 GI:21405635
 Keywords HTG; HTGS_PHASE1.
 Source human.
 Organism Homo sapiens
 Lukaryota; Metazoa; Chordata; Craniata; Vertebrates; Euteleostomi;
 Mammalia; Butheria; Primates; Catarrhini; Hominidae; Homo.
 Reference 1 (bases 1 to 93919)
 Authors Brownstein,B.H., States,D.J. and Mazzarella,R.
 Title Direct Submission
 Journal Submitted (12-AUG-1997) Center for Genomics in Medicine, Box 8232,
 Washington University School of Medicine, 4566 Scott Avenue, St.
 Louis, MO 63110, USA
 Comment On Jun 13, 2002, this sequence version replaced gi:2323256.
 Current status of this project is available at:
http://genome.wustl.edu/cgm/seq_projects.html

Submitted by:
 Ellison Chen,
 Advanced Center for Genetic Technology,
 Applied Biosystems Division of Perkin Elmer Corp.,
 850 Lincoln Center Drive,
 Foster City, CA 94404 USA
 e-mail: ellison@genseq.apldbio.com

and

Buddy Brownstein,
 Center for Genetics in Medicine, Box 8232
 Washington University School of Medicine, Box 8232
 4566 Scott Avenue,
 St. Louis, MO 63110, USA
 e-mail: buddy@genetics.wustl.edu

and

David J. States,
 Institute for Biomedical Computing
 Washington University in St. Louis
 700 South Euclid Ave.
 St. Louis, MO 63108 USA
 e-mail: states@ibc.wustl.edu.

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 6 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 9560: contig of 9560 bp in length
 * 9561 9660: gap of 100 bp
 * 9661 23227: contig of 15567 bp in length
 * 25228 25327: gap of 100 bp
 * 25328 45875: contig of 20548 bp in length
 * 45876 45976: gap of 100 bp
 * 45976 73699: contig of 27724 bp in length
 * 73700 73799: gap of 100 bp
 * 73800 82475: contig of 8676 bp in length
 * 82476 82576: gap of 100 bp
 * 82576 93919: contig of 11344 bp in length.

FEATURES

source
 ORIGIN

BASE COUNT 25991 a 21189 c 20209 g 26030 t 500 others

Query Match Score 60.6; DB 1; Length 93919;
 Best Local Similarity 79.1%; Pred. No. 0.0008; Matches 19; Mismatches 0; Indels 0; Gaps 0;
 Matches 72; Conservative 0; /clone="bwxD142"
 Qy 73 CTCCTGATATCAGTGGCGAGGAGATGACCTATGCCCTAGACAGACCTAGCG 132
 Db 982 CAGAGTGAAATAATTGGGAGGAAGTCAACATGTAGCCCTAGACAGACGAGTT 923
 Qy 133 TAGAGCTCTGTAATGATGGCTATGGCG 163
 Db 922 TACAACCTCTGAGCTGANTGGGAGTGGCG 892

Search completed: April 9, 2003, 16:53:24
 Job time : 1804 secs

Gencore version 5.1.4-p5_4578
 Copyright (C) 1993 - 2003 Compugen Ltd.

| nucleic - nucleic search, using SW model.
 run on: April 9, 2003, 15:17:04 ; Search time 1530 Seconds
 (without alignments)
 5694.884 Million cell updates/sec

Result No.	Score	Query Match Length	DB ID	Description
1	524.6	97.5	569	BG700165
2	513.0	95.4	537	BG832293
3	492.6	91.6	813	BG773070
4	492.0	91.4	527	BG1826605
5	492.0	91.4	851	BG250593
6	490.6	91.2	572	BG120838

perfect score: 538
 sequence: 1 ACGCCAGGGAGCTGTAGGC.....CTGCAAAQAAA.....AAAAA 538

scoring table: IDENTITY_NUC
 Gapop 10.0 , Gapext 1.0

searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

st-processing: Minimum Match 0%
 Maximum Match 100%

Listing first 45 summaries

database :

EST:	*
1:	em_estba:*
2:	em_estium:*
3:	em_estin:*
4:	em_estmu:*
5:	em_estov:*
6:	em_estpl:*
7:	em_estro:*
8:	em_htc:*
9:	gb_est1:*
10:	gb_est2:*
11:	gb_htc:*
12:	gb_est3:*
13:	gb_est4:*
14:	gb_est5:*
15:	em_estfun:*
16:	em_estm:*
17:	gb_gss:*
18:	em_gss_hum:*
19:	em_gss_inv:*
20:	em_gss_pln:*
21:	em_gss_yrt:*
22:	em_gss_fun:*
23:	em_gss_mam:*
24:	em_gss_mus:*
25:	em_gss_other:*
26:	em_gss_pro:*
27:	em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB	ID	Description
1	524.6	97.5	569	12	BG700165	BG700165 602679622
2	513	95.4	537	14	BM83273	BM83273 602730107
3	492.6	91.6	813	12	BG773010	BG773010 602721370
4	492	91.4	527	13	BI286605	BI286605 603077056
5	492	91.4	851	12	BG250932	BI2850932 6023638901
6	490.6	91.2	572	12	BG120838	BG120838 602351010

Toshiyuki Kuroki and Piero Carninci (RIKEN)
 Consortium (LLNL)
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLAM0704 row: 1 column: 06
 High quality sequence stop: 566.
 Location/Qualifiers

Qy	240	GGAGCTGCTCAGGACGGAGGGATGAGGACTCCACACAGACTGGCTGAGTGAAAGTGGCTCAAGGGGAAGCCTGAA	299
Db	241	GCAGCTGCTCAGGAGGATGAGGACTCCACACAGACTGGCTGAGTGAAAGTGGCTCAAGGGGAAGCCTGAA	300
Qy	300	GCTCATAGCGAGGACAGGCTCACCACAGACTGGCTGAGTGAAAGTGGCTCAAGGGGAAGCCTGAA	359
Db	301	GCTCATAGCGAGGACAGGCTCACCACAGACTGGCTGAGTGAAAGTGGCTCAAGGGGAAGCCTGAA	360
Qy	360	GGCAGGGAGGTGACCCGCCAATCAGAGGGTGAACAGGCTGAAACAGCCTGAAAGGGTGAAGAG 419	
Db	361	GGCAGGGAGGTGACCCGCCAATCAGAGGGTGAACAGGCTGAAACAGCCTGAAAGGGTGAAGAG 420	
Qy	420	CAATCACAGTGTAAAGAGACACGGTGAATGAGCACTGTCCTATGTGGAAAT 479	
Db	421	CAATCACAGTGTAAAGAGACACGGTGAATGAGCACTGTCCTATGTGGAAAT 480	
Qy	480	TGTGTCATTAAAATCTCCATTAAAGCTTTAACGGCTTCTGCAAAGAAAAAAA 536	
Db	481	TGTGTCATTAAAATCTCCATTAAAGCTTTAACGGCTTCTGCAAAGAAAAAAA 537	
RESULT 3			
ORGANISM	BG773070	813 bp mRNA sapiens cDNA clone IMAGE:4838176 5', linear	EST 15-MAY-2001
LOCUS	602721370FL NIH_MGC_97	Homo sapiens	
DEFINITION	mRNA sequence.		
REFERENCE	NIH-MGC http://mgc.ncbi.nlm.nih.gov/		
AUTHORS	Mammalian Gene Collection (MGC)		
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgsbps@mail.nih.gov		
TISSUE	Procurement: Miklos Palkovits, M.D., Ph.D. Preparation: Michael J. Brownstein (NHGRI), Shiraki		
CDDNA LIBRARY	Library Preparation: Piero Carninci (RIKEN) CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)		
DNA SEQUENCING	Sequencing by: Incyte Genomics, Inc.		
CLONE DISTRIBUTION	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov		
PLATE	LLAM1071 row: 1 column: 17		
HIGH QUALITY SEQUENCE	High quality sequence stop: 778.		
LOCATION/QUALIFIERS	Location/Qualifiers		
SOURCE	1. .813 /organism="Homo sapiens" /db_xref="Taxon:9606" /clone=IMAGE:4838176" /clone_id="NIH_MGC_97" /lab_host="DH10B"		
ORIGIN	Query Match 91.6%; Score 492.6; DB 12; Length 813; Best Local Similarity 96.8%; Pred. No. 1.8e-110; Matches 514; Conservative 0; Mismatches 14; Indels 3; Gaps 1. Base Count 215 a 163 c 209 g 226 t		

an anonymous male age 27. Library is oligo-dT primed and directionally cloned (ECOR site is destroyed upon cloning). Average insert size 1.3 kb, insert size range 0.9-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen) Research Genetics tracking code 013. Note: this is a NIH NGC Library."

Clone distribution: MGCC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at:
<http://image1.mbl.edu>
Plate: LiAM10292 row: b column: 02
High quality sequence stop: 531.
Location/Qualifiers 1 - 851

REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 572)	
AUTHORS	NIH-MGC
TITLE	http://mgc.ncbi.nlm.nih.gov/
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT	Unpublished (1999)
Contact	Robert Strausberg, Ph.D.
Email:	cgbps-r@mail.nih.gov
Tissue	Procurement: ATCC
Library Preparation:	Life Technologies, Inc.
CDNA Library Arrayed by:	The I.M.A.G.E. Consortium (LILNL)
DNA Sequencing by:	Incyte Genomics, Inc.
Clone distribution:	MGC/Clone distribution information can be found through the I.M.A.G.E. Consortium/LILNL at: http://image.llnl.gov
Plate:	LILAM10224
Row:	C
Column:	03
High quality sequence start:	5
High quality sequence stop:	572.
FEATURES	Location/Qualifiers
source	<p>1. .572</p> <p>/organism="Homo sapiens"</p> <p>/db_xref="taxon:9606"</p> <p>/clone="IMAGE:444582"</p> <p>/clone_lib="NIH MGC 90"</p> <p>/tissue_type="adenocarcinoma, cell line"</p> <p>/lab_host="DH10B (phage-resistant)"</p> <p>/note="Organ: liver; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.7 kb. Library enriched for full-length clones and constructed by Life Technologies Note: this is a NIH_MGC Library."</p>
BASE COUNT	167 a
ORIGIN	124 c 166 g 115 t

Query Match	Best Local Similarity	Score	DB 12;	Length
Best Local Matches	91.2 %	Score 490.6 ;	DB 12 ;	572 ;
Best Local Similarity	96.7 %	Pred. No.	5.5e-10 ;	2 ;
Matches	523 ;	Mleamets	14 ;	Gaps
Conservative	0 ;	Indels	4 ;	

```

/crone : IDANE :390663
/clone_id="NIH MG C_90"
/tissue_type="denoCarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/not="Organ: liver; Vector: pCMV-SPORT6; Site:1; NotI;
Site 2: SalI; Cloned unidirectionally; Oligo-dT primed.
Average insert size 1.7 kb Library enriched by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT      142 a   107 c   154 g   106 t
ORIGIN

Query Match          88.6%; Score 476.8; DB 13; Length 509;
Best Local Similarity 98.0%; Pred. No. 1..3e-106;
Matches 494; Conservatve 0; Mismatches 7; Indels 3; Gaps 1

Qy    1  AGGCCAGGGACTGTGAAATATGACTGGCAGGACTGGCTCGGACTCTTTCCCTCT 60
      6  ACGPAGGGAGCTGTGGGAGCTGTGGGAGCTGTGGCTCGGACTCTTTCCCTCT 65
Db     61 ACTGAGATTCACTGTGTAAATATGACTGGCAGGAGATCGACCC--TATGCCCTTA 117
      66 ACTGAGATTCACTGTGTAAATATGACTGGCAGGAGATCGACCTATTATGCCCTTA 125
Qy    118 GACCAAGAGCTTACGTAGGCCCTCTGAATAATGATGGGCTATGGGCCAGACTTC 177
      126 GACCAAGGGCTATGTACGCCCTCTGAATGATGGGCTATGGGCCAGACTTC 185
Db     178 GTGTGAAGTGGAAACCGAACACCTGAAAGGAACCGAACCTCAACGTCAAGTC 237
      186 GTGTGAAGTGGAAACCGAACACCTGAAAGGAACCGAACCTCAACGTCAAGTC 245
Qy    238 CTGCAGCTGCTAGGGGGAGGGATGAGGGAGCATCTGGCTGAGGGGGAAAGGCTG 297
      246 CTGCAGCTGCTAGGGGGAGGGATGAGGGAGCATCTGGCTGAGGGGGAAAGGCTG 305
Db     298 AAGCTCATGCCAGGAACGGGTCAACCCACAGACTGGGTGACTCTGAAATGGTCTGTG 357

```


/clone_lib="Soares_testis_NHT"	JOURNAL	Tumor Gene Index
/sex="male"	COMMENT	Unpublished (1997)
/lab_host="DH10B"	CONTACT	Robert Straussberg, Ph.D.
/note="Vector: pRT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo (dT) primer [5'-TGTGACCAATCTGAAGTCGGAGCGCCAATTCTTTTTTTTTTTT3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRT73 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo. "	EMAIL	cgapbsr@mail.nih.gov
BASE COUNT	101	a 139 c 105 g 144 t
ORIGIN		
Query Match	84.1%	Score 452.4; DB 9; Length 489;
Best Local Similarity	98.0%	Pred. No. 1.3e-100;
Matches	480;	Conservative 0; Mismatches 6; Indels 4; Gaps 2;
Db	43	CCGACTCTTTCTCTACTGAGATTCACTGTGTGAATAATGAGTGCGAGGAAGAT 102
Db	489	CCGACTCTTTCTCTACTGAGATTCACTGTGTGAATAATGAGTGCGAGGAAGAT 430
Qy	103	CGACCT--TATCGGCCTAGCCAAGAGCTTCTGAGCTCTGTGAATAATGAGTGGCCCTA 159
Db	429	CGACCTATTATGGCTAGCCAAGGGCTATGAGCTCTGTGAAGTGGATT-GGCCCTA 371
BASE COUNT	87	a 133 c 95 g 142 t
ORIGIN		
Query Match	80.2%	Score 431.4; DB 10;
Best Local Similarity	98.0%	Pred. No. 1.8e-95;
Matches	448;	Conservative 0; Mismatches 6; Indels 3; Gaps 1;
Qy	77	GTGAAATATGAGTGTGAGGAGGATGACCTTC--GGCCTAGACCAAGAGCTAGT 133
Db	457	GTGAAATATGAGTGTGAGGAGGATGACCTTATGGCTAGACCAAGGGCTATGT 398
Qy	134	AGAGGCTTCCTGAAATGATTGGGCCTATGGGCCGAGCAGTTCTGATGAGTGGAAACC 193
Db	397	ACAGGCTTCCTGAAATGATTGGGCCTATGGGCCGAGCAGTTCTGATGAGTGGAAACC 338
Db	337	ACGACACCTGAAAGGGGAACTCACAGTCAAGTCTGGCAGGAACTCACAGTCAAGT 278
Qy	254	GGAAGGGATGGGGACCATCTCAGGATGGGGCAAGGCTAGTGGCTATAGCCGGA 213
Db	277	GGAGGGATGGGGCATCTCAGGATGGGGCAAGGCTAGTGGCTATAGCCGGA 218
Qy	314	ACAGGGTCACCCACAGACTGGGTGAGTGGTCAAGTGGCTAGGATGGCTAGGATGGA 373
Db	217	ACAGGGTCACCCACAGACTGGGTGAGTGGTCAAGTGGCTAGGATGGCTAGGATGGA 158
Qy	374	CCGCCAAATCCAAGGGTGAACCCCTGAGAAGGTGAAGCAATCAAGCTGTTA 433
Db	157	CCGCCAAATCCAAGGGTGAACCCCTGAGAAGGTGAAGCAATCAAGCTGTTA 98
RESULT	11	
AI381509/c	LOCUS	AI381509
DEFINITION	IMAGE:2911881_x1 Soares NFL_T GBC_S1 Homo sapiens cDNA clone	447 bp mRNA linear EST 28-MAR-1999
DEFINITION	IMAGE:2911881_3' _Similar to SW.GGE4_HUMAN Q13068 GAGE-4 PROTEIN. ;	Homo sapiens CDNA clone
DEFINITION	mRNA sequence.	Similar to TR:Q13070 GAGE-6 PROTEIN. ;
ACCESSION	AV510753	TCTCCCCAATAAGCTTACAGCTTACAGCTTCTGCAAAGAAAAA 530
VERSION	AV510753.1	1
KEYWORDS		
SOURCE		
ORGANISM	Human	Human
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 457)	NCT-CGAP http://www.ncbi.nlm.nih.gov/nctgap.
AUTHORS		National Cancer Institute, Cancer Genome Anatomy Project (CGAP) ,
TITLE		

224	CACCCAGACTGGTGTGAGTGTGAAGAGGTCTGTATGGCAGAGATGGACCGCA	165							
381	AATCCAGAGGGCTGAAAACGCCATTACAGCTGAAAGGTGAAATTAAAGAAG	40							
164	AATCCAGAGGGCTGAAAACGCCATTACAGCTGAAATTAAAGAAG	105							
441	ACACGGTGAATTATGCGGGTGTCTPATGTGAAATTGTCATTAAATTCTCCA	500							
104	GCACTGTGAATTATGCGGGTGTCTPATGTGAAATTGTCATTAAATTCTCCA	45							
501	ATAAAGCTTACGCCCTTGCAAGAAAAAA	538							
44	ATAAAGCTTACGCCCTTGCAAGAAAAAA	7							
RESULT 13									
BG700088	BG700088	581 bp	mRNA	linear	EST 07-MAY-2001				
LOCUS	602679431F1	NIH_MGC_95	Homo sapiens	CDNA clone	IMAGE:4812462 5', mRNA sequence.				
DEFINITION									
ACCESSION	BG700088								
VERSION	BG700088.1								
KEYWORDS	EST.								
ORGANISM	Homo sapiens								
MATERIALS	Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.								
REFERENCE	1 (bases 1 to 581)								
AUTHORS	National Institutes of Health, Mammalian Gene Collection (MGC)								
TITLE	Unpublished (1999)								
JOURNAL	Contact: Robert Strausberg, Ph.D.								
COMMENT	Email: cgabbs-r@mail.nih.gov								
	Tissue Procurement: Miklos Palkovits, M.D., Ph.D.								
	Preparation: Michael J. Brownstein (NHGR), Shiraki Yoshiyuki and Piero Carninci (RIKEN)								
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (IIML)								
	DNA Sequencing by: Incyte Genomics, Inc.								
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/IIML at:								
	http://image.llnl.gov								
	Plate: ILMAM0704 row: m column: 07								
	High quality sequence stop: 548.								
	Location/Qualifiers								
	1 .. 581								
	/organism="Homo sapiens"								
	/db_xref=taxon:9606								
	/clone="IMAGE:4812462"								
	/clone_lib="NIH_MGC_95"								
	/tissue_type="hippocampus"								
	/lab_host="DH10B"								
	/note="Organ: brain; Vector: plvuescriptR (modified Bluescript KS+; Site 1: SalI-XbaI (gtcgag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTNTN-3', size selected for average insert size 2.5 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NHGR), National Institutes of Health. Note: this is a NIH_MGC Library."								
	164 a 142 C 166 g 109 t								
BASE COUNT									
ORIGIN									
Query Match	78.8%		Score 424;	DB 12;	Length 581;				
Best Local Similarity	92.8%		Pred. No. 1..28..93;	Indels 9;	Caps 5;				
Matches 501;	Conservative	0;	Mismatches 30;						
QY	8	GGAGCTGTGACCGAGCTGCTGTTCTGGTCCGGACTCTCTCTACTGAGA	67						
DB	25	GGAGCTGTGACCGAGCTGCTGTTCTGGTCCGGACTCTCTCTACTGAGA	84						
QY	68	TTTATCTGTGAAATTAGAGTGGCCAGGAGATCGACCTATGGCCTAGACCAAGACG	127						

NCI_CGAP_Kid5, NCI_CGAP_Kid12, NCI_CGAP_Kid3, NCI_CGAP_Kid11, NCI_CGAP_Fe12, NCI_CGAP_Co8, NCI_CGAP_Lym2, NCI_CGAP_Fe12, NCI_CGAP_Brn23, NCI_CGAP_GC4, NCI_CGAP_GC6 , NCI_CGAP_Lu24, NCI_CGAP_Lu19, NCI_CGAP_Lu21, NCI_CGAP_Brn23. These 21 libraries were pooled and a single-stranded DNA preparation of the resulting mixture was used as a tracer in a subtractive hybridization with a driver whose composition was detailed below: NCI_CGAP_Kid3 pool 1 LuAM 33343-33337, 3682-3683, 3798-3803 (IMAGE_CloneIDs 13223376-13223911, 1456008-1456775, 1500552-1502855) NCI_CGAP_Kids pool 1 LuAM 3338-3342, 3722-3725, 3776-3778 (IMAGE_CloneIDs 1322912-1325831, 1471368-1472903, 1492104-1493255) NCI_CGAP_Lu5 pool 1 LLAM pool 1 LuAM 33343-33337, 3682-3683, 3798-3803 (IMAGE_CloneIDs 1414920-1417994, 14257096-14258631, 14646064-1470983, 147592-1476743) NCI_CGAP_Pr22 pool 1 LLAM 2457-2459, 2758-2759, 3062-3065 (IMAGE_CloneIDs 9885608-988759, 1101192-1101959, 1217928-1220615) NCI_CGAP_Co1 pool 1 LLAM 2644-2872 (IMAGE_CloneIDs 1057416-1061255, 1144544-1145351). The resulting subtracted library contained 4 million recombinants. Subtraction was performed as previously described [Bonaldo et al., 1996]. Lennon & Scores (1996): Normalization and subtraction: Two Approaches To Facilitate Gene Discovery. Genome Research 6, 791-806.

TAG LIB=NCI_CGAP_GC4
TAG TISSUE=germ cell
TAG SEQ=DATTC"

74 a 119 c 87 g 138 t

SOURCE	human.
ORGANISM	Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;	
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	NIH-MGC 1 to 480 NIH-MGC http://mgc.ncbi.nih.gov/
AUTHORS	
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgbps-r@mail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, INC. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC Clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
FEATURES	High quality sequence stop: 480.
source	LocationQualifiers 1..480 /organism="Homo sapiens" /db_xref="TAXON:9606" /clone_1lib="NIH MGC 90" /tissue_type="adenocarcinoma, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: Liver; Vector: PCMV-SPORT6; Site_1: NotI; Site_2: S11; Cloned unidirectionally; oligo-dT primed. Average insert size 1.7 kb. Library enriched for

BASE COUNT	170	a	93	c	1.33	g	84	t
ORIGIN								
Query Match	76.8%		Score 413.4;	DB 1.2;	Length 480;			
Best Local Similarity	97.4%		Pred. No. 4.8e-91;					
Matches	420;	Conservative	0; Mismatches	11;	Indels 0;	Gaps		
Qy	108	TATGGCCCTAGACCAAGAACGCTACTGAGCTCTGAATCATGGCCCTATGCCGCC 167						
Db	15	TATGGCCCTAGACCAAGAACGCTACTGAGCTCTGAATCATGGCCCTATGCCGCC 74						
Qy	168	GAGCAGTTCACTGTGAACTGGAAACCAGAACACCTGAAAGAAGGGGAAACGAACTCAA 227						
Db	75	GAGCAGTTCACTGTGAACTGGAAACCAGAACACCTGAAAGAAGGGGAAACGAACTCAA 134						
Qy	228	CCTCAGGTTACCTGAGCTCTCGAGGCTCTCGAGGGAGAGATGGGAGACATCTGAGTCAGGC 287						
Db	135	CGTCAGGATCTCTGAGCTCTCGAGGCTCTCGAGGGAGAGATGGGAGACATCTGAGTCAGGC 194						
Qy	288	CGGAAGGCTGAAAGTCATGCCAGGAACACGGTCACCCACAGACTGGGTGAGTGTRGAA 347						
Db	195	CGGAAGGCTGAAAGTCATGCCAGGAACACGGTCACCCACAGACTGGGTGAGTGTRGAA 254						
Qy	348	GATGGTCCTGATGGCAAGAGATGGACCCGCAAATCCAGGGAGTGAAGACGCCGCTGAA 407						
Db	255	GATGGTCCTGATGGCAAGAGATGGACCCGCAAATCCAGGGAGTGAAGACGCCGCTGAA 314						
Qy	408	GAAGGTGAAAGGAATCACAGTGTAAAGGAAGACRGTTGAAATGTCGGGTGCTCC 467						
Db	315	GAAGGTGAAAGGAATCACAGTGTAAAGGAAGGCACTGCAATGTCGGGTGCTCC 374						
Qy	468	TATGTTGCAATTGTTCTTTAAATTCTCCCAATAAGCTTACAGCTCTCTGCAAAAGA 527						
Db	375	TATGTTGCAATTGTTCTTTAAATTCTCCCAATAAGCTTACAGCTCTCTGCAAAAGA 434						
Qy	528	AAAAAAAAAA 538						
Db	435	AAAAAAAAAA 445						

Search completed: April 9, 2003; 17:06:32

Wed Apr 9 18:05:23 2003

us-09-782-745-14.rst

Page 11

Job time : 1535 secs



XX 13-DEC-2001.
 PD 2000US-209473P.
 XX 30-MAY-2001; 2001IWO-US10838.
 PR 05-JUN-2000; 2000US-209531P.
 PR 18-SEP-2000; 2000US-23313P.
 PR 20-SEP-2000; 2000US-234005P.
 PR 20-SEP-2000; 2000US-234034P.
 PR 22-SEP-2000; 2000US-234052P.
 PR 22-SEP-2000; 2000US-234567P.
 PR 25-SEP-2000; 2000US-234922P.
 PR 25-SEP-2000; 2000US-234924P.
 PR 25-SEP-2000; 2000US-235077P.
 PR 25-SEP-2000; 2000US-235082P.
 PR 25-SEP-2000; 2000US-235134P.
 PR 25-SEP-2000; 2000US-235280P.
 PR 26-SEP-2000; 2000US-235631P.
 PR 26-SEP-2000; 2000US-235633P.
 PR 27-SEP-2000; 2000US-235711P.
 PR 27-SEP-2000; 2000US-235720P.
 PR 27-SEP-2000; 2000US-235844P.
 PR 28-SEP-2000; 2000US-235863P.
 PR 28-SEP-2000; 2000US-236028P.
 PR 28-SEP-2000; 2000US-236032P.
 PR 28-SEP-2000; 2000US-236034P.
 PR 28-SEP-2000; 2000US-236109P.
 PR 28-SEP-2000; 2000US-236111P.
 PR 29-SEP-2000; 2000US-236842P.
 PR 29-SEP-2000; 2000US-236891P.
 PR 02-OCT-2000; 2000US-237172P.
 PR 02-OCT-2000; 2000US-237173P.
 PR 02-OCT-2000; 2000US-237278P.
 PR 02-OCT-2000; 2000US-237294P.
 PR 02-OCT-2000; 2000US-237295P.
 PR 03-OCT-2000; 2000US-237316P.
 PR 03-OCT-2000; 2000US-237425P.
 PR 03-OCT-2000; 2000US-237598P.
 PR 03-OCT-2000; 2000US-237604P.
 PR 03-OCT-2000; 2000US-237606P.
 PR 01-NOV-2000; 2000US-237608P.
 PR 01-NOV-2000; 2000US-244861P.
 PR 01-NOV-2000; 2000US-245084P.
 XX (AVAL-) AVALON PHARM.

XX Young PE, Augustus M., Carter KC, Ebner R, Endress G, Horrigan S;
 PI Soper DR, Weaver Z;
 XX WPI: 2002-108264/24.

PT Screening for anti-neoplastic agent involves exposing cells to a
 PT chemical agent to be tested for anti-neoplastic activity, and
 PT determining a change in expression of a gene of a signature gene set -

XX Claim 1; SEQ ID 4660; 44pp; English.

CC The present invention describes a method (M1) for screening for an
 CC anti-neoplastic agent. The method involves exposing cells to a chemical
 CC agent to be tested for anti-neoplastic activity, determining a change in
 CC expression of at least one gene (I) of a signature gene set, where (I)
 CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
 CC to AB170110), or is at least 95% identical to (S), where a change in
 CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
 CC activity and can be used in gene therapy. M1 can be used for screening
 CC an anti-neoplastic agent, and can be used for producing a product which
 CC is the data collected with respect to the anti-neoplastic agent as a
 CC result of M1, and the data is sufficient to convey the chemical as a
 CC structure and/or properties of the agent. M1 can be used in the

CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
 CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
 CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
 CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
 CC carcinoma, papillary carcinoma and Wilm's tumour.
 XX SQ Sequence 528 BP; 147 A; 114 C; 156 G; 111 T; 0 other;
 XX SQ Score 501.4%; DB 24; Length 528;
 XX Best Local Similarity 93.2%; Pred. No. 1.7e-135;
 XX Matches 518; Conservative 0; Mismatches 6; Indels 3; Gaps 1;

Qy	2 CGCAGGGAGGTGAGGCAAGTGTGTTCTGGCTCGGACTCTTTTCCTCTTA 61
Ds	1 CGCAGGGAGGTGAGGCAAGTGTGTTCTGGCTCGGACTCTTTTCCTCTTA 60
Qy	62 CTGAGATTCACTGTGTGAATAATGAGTGGGAGAAATGAGC---TATGGGCTAG 118
Ds	61 CTGAGATTCACTGTGTGAATAATGAGTGGGAGAAATGAGCTTAG 120
Qy	119 ACCAAGACGCTACGTAGGCCCTATGGCCCTAGGCCAGCAGTTCAG 178
Ds	121 ACCAAGCGCTATGTCAGCCCTCTGAATGATGGCCTATGGCTATGGCAGCAGTCAG 180
Qy	179 TGATGAACTGGAAACCAGAACACTGAAGAAGGGAAACCACTCAACGTCAGATCC 238
Ds	181 TGATGAACTGGAAACCAGAACCTGAAGAAGGGAAACCACTCAACGTCAGATCC 240
Qy	239 TGCAGCTGCTCAGGAGGAGGATGAGGAGCATCTGAGGTCAAGGGCGAAGGCTGA 298
Ds	241 TGCAGCTGCTCAGGAGGAGGATGAGGAGCATCTGAGGTCAAGGGCGAAGGCTGA 300
Qy	299 AGCTCAATGCCAGGAAACAGGGTACCCACAGACTGGGTGAGATGGTCTCTGA 358
Ds	301 AGCTGATGCCAGGAAACAGGGTACCCACAGACTGGGTGAGATGGTCTCTGA 360
Qy	359 TCGCAGGAGATGGTACCCACAGACTGGGTGAGATGGTCTCTGA 418
Ds	361 TCGCAGGAGATGGTACCCACAGACTGGGTGAGATGGTCTCTGA 420
Qy	419 GCATCACAGTGTAAAGAAGACAGCTGAAATGAGCAGGTGCTCTATGTTGGAAA 478
Ds	421 GCATCACAGTGTAAAGAAGGAGCTGAAATGAGCAGGTGCTCTATGTTGGAAA 480
Qy	479 TTGTTCTTCAATTAATTCCTCCAAATAAGCTTACGCCCTCTGGCAA 525
Ds	481 TTGTTCTTCAATTAATTCCTCCAAATAAGCTTACGCCCTCTGGCAA 527

RESULT 7
 AA18721
 ID AA18721 standard; CDDNA; 539 BP.
 XX AA18721;
 AC
 XX DT 30-JUL-1998 (first entry)
 XX CDDNA encoding GAGE-6 tumour rejection antigen precursor.
 XX GAGE tumour rejection antigen precursor; TRAP; tumour;
 KW diagnosis; melanoma; antigen; cytolytic T cell clone proliferation;
 KW HLA-typing assay; 88.
 XX Homo sapiens.
 OS
 XX Key
 FH 82..436
 FT /*tag= a
 FT /transl_except= (pos:127..129, aa:Arg)
 FT /transl_except= (pos:196..198, aa:Ala)
 FT /transl_except= (pos:199..201, aa:Thr)
 XX PN WO9749417-A1.

DT	30-JUL-1998	(first entry)	
XX	DE	CDNA encoding GAGE-4 tumour rejection antigen precursor.	
XX	KW	GAGE tumour rejection antigen precursor; TRAP; tumour diagnosis; melanoma; antigen; cytolytic T cell clone proliferation; HLA-typing assay; B8.	
XX	KW	Homo sapiens.	
OS	XX	Key CDS	Location/Qualifiers 83..435 /*tag= a /transl_except= (pos:197..199, aa:Ala) /transl_except= (pos:200..202, aa:Thr)
PN	XX	WO9749417-A1.	
PH	XX	PD 31-DEC-1997.	
FT	XX	PF 23-JUN-1997;	
FT	XX	PR 24-JUN-1996;	
FT	XX	PA (LUDW-) LUDWIG INST CANCER RES.	
PI	XX	WPI; 1998-076905/07.	
DR	XX	P-FSDB; AAW47601.	
PT	XX	Isolated nucleic acid encoding GAGE tumour rejection antigen precursor - processed by HLA-Cw6 molecules into peptides, useful to diagnose melanomas	
PT	XX	PS Example 13: Fig 4; 60pp; English.	
CC	XX	The present sequence encodes a GAGE-4 tumour tumour rejection antigen precursor (TRAP). The protein is expressed in a number of tumours. In contrast the only normal tissue which expresses GAGE TRAP protein is testis. Several GAGE TRAPS have been identified (see AAV18717-21). The major difference between these proteins and GAGE-1 (AAV05540) is the absence of a stretch of 143 bases located at position 379 to 521 of the GAGE-1 TRAP sequence. The rest of the sequences show mismatches at various position, with the exception of GAGE-3 whose 5' end is totally different from the other GAGE cDNAs for the first 112 bases. This region of GAGE-3 cDNA contains a long repeat and a hairpin structure. The antigens can be used to diagnose melanomas, characterised by expression of a TRAP or presentation of a tumour rejection antigen. Antigens shed into blood or urine can be observed and then used to confirm a diagnosis of melanoma using cytolytic T cell clone proliferation methodologies. Other uses for the processed peptides, include HLA-typing assays for, e.g. skin graft or organ transplants.	
SQ	XX	Sequence 541 BP; 158 A; 117 C; 154 G; 112 T; 0 other;	
Query Match	91.2%	Score 490.6; DB 19; Length 541;	
Best Local Similarity	96.7%	Pre. No. 2..3e-132;	
Matches 523;	Conservative 0;	Mismatches 14;	
		Indels 4; Gaps 2;	
Qy	2	CGCAGGGAGCTGTCAGGGCACTGCTGTGCTGCCTGCCTGGCGACCTTTCCTCTCA 61	
Db	1	CGCACAGGAGCTGTCAGGGCACTGCTGTGCTGCCTGCCTGGCGACCTTTCCTCTCA 60	
Qy	62	CAGGATCATCTGCTGAAATAATGAGTTGGGAGGAGATCACC--TATCGGCCCTAG 118	
Db	61	CAGGATCATCTGCTGAAATAATGAGTTGGGAGGAGATCACCCTATATGGCCCTAG 120	
Qy	119	ACAGAGGCTACCTAGAGCCTCTGAATGATGGGCTATGGCCGAGGAGTTCA 178	
Db	121	ACAGGCTATGCTCAGCTCTGAATGATGGCCCTATGGCCGAGGAGTTCA 180	
Qy	179	TGATGAAGTGGAAACCGCAACCTGAGAAAGGGAAACCGCAACTCAACGTCAAGGATCC 238	
Db	181	TGATGAAGTGGAAACCGCAACCTGAGAAAGGGAAACCGCAACTCAACGTCAAGGATCC 240	
Qy	239	TGAGCTCTGAGGGAGGATGGGGCATCTGAGGTCAAGGGCAACCTGA 298	
Db	241	TGAGCTCTGAGGGAGGATGGGGCATCTGAGGTCAAGGGCAACCTGA 300	
Qy	299	AGCTCATAGGCCAGAACAGGGTCAACCACAGACTGGGTGAGATGCTCTGCTGA 358	
Db	301	AGCTCATAGGCCAGAACAGGGTCAACCACAGACTGGGTGAGATGCTCTGCTGA 360	
Qy	359	TGGGCAGGAGATGGCCCAAATCAGAGGGTGAAGAGCTGAGAGTGGAAA 418	
Db	361	TGGCAGGAGATGGACCCCAAATCAGAGGGTGAAGAGCTGAGAGTGGAAA 420	
Qy	419	GCAATCCAGTGTAAAGAAGAACGTTGAAATGAGGCTGCTCTATGTGGAA 477	
Db	421	GCAATCCAGTGTAAAGAAGAACGTTGAAATGAGGCTGCTCTATGTGGAA 480	
Qy	478	ATTGGTCATTAAATCTCCCAAATAAGCTTACGGCCTTCTGAAGAAAAAAA 537	
Db	481	ATTGGTCATTAAATCTCCCAAATAAGCTTACGGCCTTCTGAAGAAAAAAA 540	
Qy	538	A 538	
Db	541	A 541	
RESULT	11		
AAX90520	ID	AAX90520 standard; cDNA; 560 BP.	
XX	XX	AAX90520;	
AC	XX	DT 30-SEP-1999 (first entry)	
XX	XX	DE GAGE-3 tumour rejection antigen clone nucleotide sequence.	
XX	XX	Human leukocyte antigen; HLA-A29; tumour rejection antigen; detection; therapy; pathological condition; cancer; CTL; cytolytic T lymphocyte; GAGE; 58.	
XX	XX	Hom sapiens.	
XX	XX	W0937665-A1.	
XX	XX	29-JUL-1999.	
PD	XX	PP 12-JAN-1999 99WO-US00775.	
XX	XX	23-JAN-1998; 98US-0012818.	
PA	XX	(LUDW-) LUDWIG INST CANCER RES.	
PI	XX	Boon-Falleur T, Debacker O, Van Den Eynde B, Van Der Bruggen P;	
XX	XX	WPI; 1999-469111/39.	
PT	XX	New isolated peptides which bind to HLA-A29 molecules, which are tumour rejection antigens used for detection and therapy of pathological conditions, e.g. cancer	
PT	XX	Example 13; Fig 4; 62pp; English.	
CC	XX	The present invention describes peptides which bind to human leukocytes antigen (HLA) A29 (HLA-A29) molecules. The peptides are processed into tumour rejection antigens. They can be used for detecting cytotoxic T lymphocytes (CTLs) in pathological conditions such as cancer and in HLA-typing assays. Complexes of HLA-A29 molecules and the peptides can be used for stimulating CTLs in vivo. The present sequence represents a GAGE tumour rejection antigen clone, from an example from the present invention.	

SQ	Sequence 560 BP; 177 A; 119 C; 150 G; 114 T; 0 other;	
	Query Match 97.9%; Score 429.2; DB 20; Length 560;	
	Best Local Similarity 92.2%; Pred. No. 1; 6e-114;	
	Matches 452; Conservative 0; Mismatches 38; Indels 0; Gaps 0;	
Qy	49 TCTTTTCTCTACTCTGATTCATCTGTGAATAATAGTGGCAGGAAGATCGACCT 108	
Db	71 TCTTCTCGCACTATTCACAGATAATCTAGAGAAATCCACCTATT 130	
Qy	109 ATCGGCCTAGACCAAGCTGCTACGTAGAGCCCTATGGGCCG 168	
Db	131 ATGGCTGAGCCAAGCGCTATGGCTCTGAGTGATGGCCCTATGGCCG 190	
Qy	169 AGCACTTCACTGATGAGTGGANCCAGAACCTGAGAAGGGAAACCTCACAC 228	
Db	191 AGCACTTCACTGATGAGTGGACCAACCTGAGAAGGGAAACCTCACAC 250	
Qy	229 GTCAAGATCTCTGAGCTGCTAGGTGCTAGGTCAAGGC 288	
Db	251 GTCAAGATCTCTGAGCTGCTAGGTGCTAGGTCAAGGC 310	
Qy	289 GAAACCTGAGCTATAGCAGGTAACCCAGACTGGGTGAGTGTGAG 348	
Db	311 CGAACCTGAGCTGATGCCAGACGGTCAACCCAGACTGGGTGAGTGTGAG 370	
Qy	349 ATGGTCTGATGGCAGGAGATGGACGGGAAACGCTGAG 408	
Db	371 ATGGTCTGATGGCAGGAGATGGACGGGAAACGCTGAG 430	
Qy	409 AAGGTGAAAGCAATCACAGTGTAAAGAAGACCTGTAAATGATGTCAGGTCTCT 468	
Db	431 AAGGTGAAAGCAATCACAGTGTAAAGAAGACCTGTAAATGATGTCAGGTCTCT 490	
Qy	469 ATGTTGAAATTGTTCTTAAATTCTCCAAATAAACCTTACAGCTTCTGAAAGAA 528	
Db	491 ATGTTGAAATTGTTCTTAAATTCTCCAAATAAACCTTACAGCTTCTGAAAGAA 550	
Qy	529 AAAA..... 538	
Db	551 AAAA..... 560	
RESULT 12		
ID	AAV18718 standard; cDNA; 560 BP.	
XX		
AC	AAV18718;	
XX		
DT	3-0-JUL-1998 (first entry)	
XX		
DE	cDNA encoding GAGE-3 tumour rejection antigen precursor.	
XX		
KW	GAGE tumour rejection antigen precursor; TRAP; tumour;	
KW	diagnosis; melanoma; antigen; cytolytic T cell clone proliferation;	
KW	HLA-typing assay; ss.	
XX	Homo sapiens.	
XX	Location/Qualifiers	
FH	100..457	
FT	/tag= a	
FT	/transl_except= {pos:217..219, aa:Ala}	
FT	/transl_except= {pos:220..222, aa:Thr}	
XX	WO9749417-A1.	
XX	31-DEC-1997.	
PF	23-JUN-1997;	97WO-US10850.
XX		
PR	24-JUN-1996;	96US-0669161.
XX		

(LUDW-) LUDWIG INST CANCER RES.
 PA XX Boon-Pailleur T, Debacker O, Van Den Eynde B;
 PI XX WPI: 1998-076905/07.
 DR XX P-PSDB; AAW7600.
 PT Isolated nucleic acid encoding GAGE tumour rejection antigen precursor - processed by HLA-Cw6 molecules into peptides, useful to diagnose melanomas
 PT XX Example 13; Fig 4; 60PP; English.
 PS XX
 XX The present sequence encodes a GAGE-3 tumour tumour rejection antigen precursor (TRAP). The protein is expressed in a number of tumours. In contrast the only normal tissue which expresses GAGE TRAP protein is testis. Several GAGE TRAPS have been identified (see AAV18718-1). The major difference between these proteins and GAGE-1 (AAV05540) is the absence of a stretch of 143 bases located at position 319 to 521 of the GAGE-1 TRAP sequence. The rest of the sequences show mlematches at various positions, with the exception of GAGE-3 whose 5' end is totally different from the other GAGE cDNAs for the first 112 bases. This region of GAGE-3 cDNA contains a long repeat and a hairpin structure. The antigens can be used to diagnose melanomas, characterised by expression of a TRAP or presentation of a tumour rejection antigen. Antigens shed into blood or urine can be observed and then used to confirm a diagnosis of melanoma using cytotoxic T cell clone proliferation methodologies. Other uses for the processed peptides, include HLA-typing assays for, e.g. skin graft or organ transplants.
 XX SQ Sequence 560 BP; 175 A; 119 C; 152 G; 114 T; 0 other;
 SQ Query Match 78.3%; Score 421.2; DB 19; Length 560;
 SQ Best Local Similarity 91.2%; Pred. No. 3.4e-112;
 SQ Matches 447; Conservative 0; Mlematches 43; Indels 0; Gaps 0;
 Qy 49 TCTTTTCTCTCTACTGAGTTCATCTGTGTGAATAATGAGTTGGCAGGAAAGATCGACCT 108
 Db 71 TCGTCCTCGCCAACTCATATTTCACAGATGAACTCTCTGAGGAAAATGACCTATT 130
 Qy 109 ATGGCCCTAGACCAAGACGCTACTAGAGCTCTAGAGCTCTCTGAAATCTGGCCCTATTCGGCCCG 168
 Db 131 ATGGCCCTAGACCAAGGCTATGAGCTCTGAGTGTAAATGAGTGTAAATGAGCTCTATGGCCCTATTCGGCCCG 190
 Qy 169 AGCAAGTTCAGTTGAACTGGAAACCTGAAACACCTGAAAGAGGGAAACCACTAAC 228
 Db 191 AGCAAGTTCAGTTGAACTGGAAACAGGCTCTGAAAGGGAAACGACTAAC 250
 Qy 229 GTCAGGATCTGAGCTGCTCAGGGAGAGATGAGGAGCATCTGAGGTCAAGGGC 288
 Db 251 GTCAGGATCTGAGCTGCTCAGGTGAGGAGCATCTGAGGTCAAGGGC 310
 Qy 289 CGAGGCTCTGAGGTCTATCCAGGAACGGGTACCCAGACTGGGTGAGTGTGAG 348
 Db 311 CGAGGCTCTGAGGTCTATCCAGGAACGGGTACCCAGAGCTGGGTGAG 370
 Qy 349 ATGGTCCCTGATGGCAGGATGGACCCGCAAAATCAGAGGTTAAAAGCTGTGAG 408
 Db 371 ATGGTCCCTGATGGCAGGATGGACCCGCAAAATCAGAGGTTAAAAGCTGTGAG 430
 Qy 409 AAGGTGAAAGCAATCACAGTGTAAAGAAGACCTGAAATGATGTCAGGTCTCT 468
 Db 431 AAGGTGAAAGCAATCACAGTGTAAAGAAGGCAAGGTGAAATGATGTCAGGTCTCT 490
 Qy 469 ATGGTCCCTGATGGCAGGATGGACCCGCAAAATCAGAGGTTAAAAGCTGTGAG 528
 Db 491 ATGGTCCCTGATGGCAGGATGGACCCGCAAAATCAGAGGTTAAAAGCTGTGAG 550
 Qy 529 AAAA..... 538
 Db 551 AAAA..... 560

RESULT 13	Qy	370 TGGACCCGCAATCCGAGGGCTGTGAAAGGTGCTGAAAGCTCATCAGT 429
ID AAZ97217	Db	340 TGGACCCGCAATCCGAGGGCTGTGAAAGGTGCTGAAAGCTCATCAGT 399
XX	Qy	430 GTTAAAAGAACGACCGTGTGAAATTGTCATTTGTCATTA 489
AC	Db	400 GTTAAAAGAACGACCGTGTGAAATTGTCATTTGTCATTA 459
XX	XX	18-APR-2000 (first entry)
DE	XX	Human prostate cancer differentially expressed gene #78.
XX	XX	Prostate cancer specific gene; cancer; tumour progression; diagnose;
XX	XX	hyperproliferative cell growth; prostatic disorder; treatment;
KW	XX	metastatic prostate cancer; benign prostate hyperplasia; BPH; ss.
KW	XX	
OS	XX	RESULT 14
XX	XX	AAZ97216/c
PN	XX	AAZ97216 standard; cDNA; 1024 BP.
XX	XX	ID AAZ97216;
XX	XX	AC AAZ97216;
PD	XX	18-APR-2000 (first entry)
PF	XX	XX
PR	XX	Human prostate cancer differentially expressed gene #77.
PR	XX	XX
PR	XX	Prostate cancer specific gene; cancer; tumour progression; diagnose;
PR	XX	hyperproliferative cell growth; prostatic disorder; treatment;
PA	XX	KW metastatic prostate cancer; benign prostate hyperplasia; BPH; ss.
PA	XX	
PI	XX	XX
PI	XX	Homo sapiens.
PI	XX	OS Homo sapiens.
DR	XX	XX
DR	XX	PN WO9964594-A2.
DR	XX	XX
PT	XX	16-DEC-1999.
PT	XX	XX
PT	XX	99WO-US13181.
PT	XX	XX
PS	XX	10-JUN-1999;
PS	XX	XX
PS	XX	98US-0088877.
PS	XX	XX
PS	XX	09-JUN-1999;
PS	XX	XX
PS	XX	99US-0088877.
PS	XX	XX
PS	XX	PR 11-JUN-1998;
PS	XX	XX
PS	XX	98US-0088877.
PS	XX	XX
PS	XX	09-JUN-1999;
PS	XX	XX
PS	XX	(CHIR) CHIRON CORP.
PS	XX	XX
PS	XX	Astel JH, Carroll E, Endege WO, Ford DM, Monahan JE, Schlegel R;
PS	XX	Steinmann KE, Zhang J;
PS	XX	XX
PS	XX	WPI; 2000-116541/10.
PS	XX	XX
PS	XX	New isolated prostate cancer specific nucleic acids, used to develop
PS	XX	products for the diagnosis and treatment of cancer -
PS	XX	Claim 2 ; Page 107; 212pp; English.
CC	XX	This sequence represents a prostate cancer specific nucleic acid
CC	XX	sequence. The invention relates to a method for diagnosing cancer,
CC	XX	tumour progression, hyperproliferative cell growth or accompanying
CC	XX	biological and physical manifestations. The method involves contacting
CC	XX	the biological sample with a probe that comprises a sequence capable of
CC	XX	hybridising to any of the 339 nucleotide sequences given in the
CC	XX	specification (see AAZ97140-297478) and detecting duplex formation. The
CC	XX	products and methods of the invention can be used for the diagnosis,
CC	XX	prognosis, and treatment of cancer, tumour progression,
CC	XX	hyperproliferative cell growth, and accompanying physical and biological
CC	XX	manifestations. They can be used particularly for prostatic disorders
CC	XX	such as metastatic prostate cancer, localised prostate cancer, or benign
CC	XX	prostate hyperplasia (BPH).
CC	XX	Sequence 1024 BP; 272 A; 243 C; 275 G; 207 T; 27 other;
SQ	XX	Query Match Score 74.2%; Score 399.4%; DB 21; Length 1024;
Best Local Similarity 98.5%; Pred. No. 9 6e-106; Indels 0; Gaps 0;	Matches 403; Conservative 0; Mismatches 6;	
Qy 130 AGTAGACCCCTCGAAATGATGATGGCCATGCGCCGAGGTTAGTGAAGTGC 189		
Db 100 AGGTACGCTCTGGAAATGATGCGCCATGCGCCGAGGTTAGTGAAGTGC 159		
Qy 190 AACCGAAACACCTGAAGAGGGGACCCGAACTCAACGCTGAGCTGTC 249		
Db 160 AACAGAACACCGAAGGGGAACTCAACGCTGAGCTGTC 219		
Qy 250 AGGAGGGAGGAGGATGAGGGACATTCGCACTGGCTGAAAGCTCATAGC 309		
Db 220 AGGAGGGAGGAGGATGAGGGACATTCGCACTGGCTGAAAGCTCATAGC 279		
Qy 310 AGGAACAGGGTCAACCCACAGACTGGTGTAGTGTGAAAGATGCTCTGATGCCAGGACA 369		
Db 280 AGGAACAGGGTCAACCCACAGACTGGTGTAGTGTGAAAGATGCTCTGATGCCAGGACA 339		
Qy	Query Match Score 73.9%; Score 397.4%; DB 21; Length 1024;	
Best Local Similarity 98.0%; Pred. No. 3.6e-105; Indels 0; Gaps	Matches 401; Conservative 0; Mismatches 8;	
Qy 130 ACGTAGAGGCCCTCGAAATGATGATGGCCATGCGCCGAGGTTAGTGAAGTGC 189		
Db 521 AGCTTGCGCTCCCTGAAATGATGATGGCCATGCGCCGAGGTTAGTGAAGTGC 462		

Key CDS	Location/Qualifiers 49 . 469 /*tag = a /note= "putative open reading frame"
---------	--------------------------------------------------------------------------------------

WO9621673-A1:

18-JUL-1996. 96WO-US0031
11-JAN-1996; 95US-05316
21-SEP-1995; 95US-03706
10-JAN-1995; {LUDWIG- LUDWIG INST CANCE

Boon-Falleur T., Debacker O., Van Den Eynde B., Van Der Bruggen P.; WPI: 1996-342237/34.

Isolated tumour rejection antigen genes and peptide(s) - used to develop Prods. for diagnosis and treatment of abnormal conditions e.g. cancer

Business Case Name:

The present sequence is GAGE tumour rejection antigen (TRA) precursor (TRAP) gene, derived from the human melanoma cell line PCMV-US92/04351. It encodes peptides that are processed into human MHC HLAB29 mol. binding peptides i.e. TRA.



GenCore version 5.1.4_p5_4578
 Copyright (c) 1993 - 2003 Compugen Ltd.

MM protein - protein search, using sw model

 run on: April 7, 2003, 18:13:10 ; Search time 43 Seconds
 (without alignments)
 259.339 Million cell updates/sec

 title: US-09-782-745-27
 perfect score: 646
 sequence: I M S H R G R S T Y R P R R Y V E P P D P P N P E E V K T P E E G E K Q S Q C 116
 scoring table: BLOSUM62
 Gapcost 10.0 Gapext 0.5

searched: 28324 seqs, 9613422 residues

THE JOURNAL OF CLIMATE

卷之三

Maximum Match 100%

Digitized by srujanika@gmail.com

2: **pir2:***

314 *Journal of Health Politics, Policy and Law*

No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, derived by random sampling from the population.

Result No.	Score	Query Match		Length	DB ID	Description
		Length	DB			
1	119	18.4	1300	2	T03166	probable glutamyl proline-rich surface protein
2	110	17.0	139	2	S70010	glutamyl proline-rich surface protein
3	110	17.0	221	2	S70019	glutamyl proline-rich surface protein
4	106	16.4	301	2	E29149	glutamyl proline-rich surface protein
5	105.5	16.3	260	2	S22373	glutamyl proline-rich surface protein
6	104	16.1	396	1	A58338	glutamyl proline-rich surface protein
7	103.5	16.0	171	2	A27077	glutamyl proline-rich surface protein
8	102.5	15.9	815	2	B30043	glutamyl proline-rich surface protein
9	102.5	15.9	815	2	JN0689	glutamyl proline-rich surface protein
10	102.5	15.9	830	2	S15720	glutamyl proline-rich surface protein
11	102.5	15.9	848	2	S02262	glutamyl proline-rich surface protein
12	102	15.8	652	2	PC4114	calcium binding protein
13	102	15.8	1560	2	T30882	proline-rich protein
14	101.5	15.7	300	2	S19860	hypothetical protein
15	101	15.6	754	2	B97386	OmpA family protein
16	101	15.6	754	2	AC2007	beta-CCP domain containing protein
17	100.5	15.6	605	1	PW58BA	hypothetical protein
18	100.5	15.6	847	2	F96331	zinc metalloprotease inhibitor
19	100	15.5	1881	2	H95576	proline-rich protein
20	99.5	15.4	309	2	S10089	neurofilament protein
21	99.5	15.4	548	1	QFPGL	beta-CCP domain containing protein
22	99.5	15.4	605	2	S20007	NF-180 protein
23	99.5	15.4	1110	2	I51116	proline-rich protein
24	98	15.2	212	2	B36298	glutamyl proline-rich surface protein
25	98	15.2	838	1	EEWHW	paranormal protein
26	98	15.2	1606	2	T34073	hypothetical protein
27	98	15.2	2649	2	T5123	salivary protein
28	97.5	15.1	166	2	B25372	hypothetical protein
29	97.5	15.1	3436	2	S55559	hypothetical protein

Query Match Score DB 2; Length 1300;

586 DEGPEGFPGPEGCPGGPEGECPGGPEGLEPFGDEGPEEGG5EGDSDPDGGAQEVPEGPKG 645
107 PEECEKOS 114

646 P-BGECQS 652

0010 - **Ucutanate/Proline-rich Protein (clone BB, 1.4.1) - rat**
Species: *Rattus norvegicus* (Norway rat)

Accession S70010 Geochman, R.; McMahon, A.; Sabban, E.L.
Geochman. Biophys. Acta 1306, 147-152, 1996

A case-control study of risk factors for primary liver cancer in Taiwan

Roots: 1-139 <GEE>
Residues: U-139
Cross-references: EMBL:U40628; NID:91184695; PIDN:AAB05668.1; PID:91184696

Matches 34; Conservative 15; Mismatches 36; Indels 22; Gaps 5;

Qy 15 RYVPEPMIGPMRPEQFSDEVEPATPEEGE -- -PATORQDPAAAQEGEDEGASAGQGPKP 71
 Db 47 RASEEEETGGP --- EDSEAVQPRTEPEPELEAATKGDEPDLQEMEPF ----- PEP 95

Qy 72 EAHSQEQQHPTQGCECEDGDQEMDPPNPEEVKTP - EGEKQSQC 116
 Db 96 EVPEPEPEPEPEPEPEPEB ----- PEPPEREDFQEGDESEGK 136

RESULT 3

Glutamate/proline-rich protein (clone 53.1.1.1) - rat
 C;Species: Rattus norvegicus (Norway rat)
 C;Accession: S70009
 R;Geerman, R.; McMahon, A.; Sabban, E.L.
 R;Biochim. Biophys. Acta 1306, 147-152, 1996
 A;Title: Cloning and characterization of cDNAs for novel proteins with glutamic acid-pro
 A;Reference number: S70009; MUID:962183; PMID:863331
 R;Accession: S70009
 A;Status: preliminary
 A;Molecule type: mRNA
 R;Residues: 1-221 <GEE>
 R;Cross-references: EMBL:U40627; NID:gi184693; PIDN:AAB05667.1; PID:91184694

Query Match Score 110; DB 2; Length 221;
 Best Local Similarity 31.8%; Pred. No. 0.16;
 Matches 34; Conservative 15; Mismatches 36; Indels 22; Gaps 5;

Qy 15 RYVPEPMIGPMRPEQFSDEVEPATPEEGE -- -PATORQDPAAAQEGEDEGASAGQGPKP 71
 Db 129 RASEEEETGGP --- EDSEAVQPRTEPEPELEAATKGDEPDLQEMEPF ----- PEP 177

Qy 72 EAHSQEQQHPTQGCECEDGDQEMDPPNPEEVKTP - EGEKQSQC 116
 Db 178 EVPEPEPEPEPEPEPEPEB ----- PEPPEREDFQEGDESEGK 218

RESULT 4

B29149 Proline-rich protein - mouse
 C;Species: Mus musculus (house mouse)
 C;Date: 08-Mar-1989 #sequence_revision 08-Mar-1989 #text_change 20-Aug-1999
 C;Accession: E29149
 R;Clements, S.; Mehansho, H.; Carlson, D.M.
 J. Biol. Chem. 260, 13471-13477, 1985
 A;Title: Novel multigene families encoding highly repetitive peptide sequences. Sequence
 A;Reference number: A92201; MUID:86033799; PMID:3840480
 A;Content: Clone PMP125
 A;Accession: E29149
 A;Molecule type: mRNA
 A;Residues: 1-301 <CLE>
 C;Superfamily: proline-rich protein

Query Match Score 106; DB 2; Length 301;
 Best Local Similarity 29.3%; Pred. No. 0.43;
 Matches 39; Conservative 10; Mismatches 54; Indels 30; Gaps 7;

Qy 4 RGRSTYR - PRRRYVPEPMIGP-MRPEQFSDEVEPATPEEGEPATQDPAAAQEGEDE 61
 Db 121 QSSQQRPQPGRQQGPPQQGPQRPPQPGNQGPPQ 178

Qy 62 GASAG --- QGPKEPAHSQ --- EGHHDTGCGCEDGPGQ --- QEMDPNP 100
 Db 179 CGPQGPGRPNQGNGQPPGQGPQRTQPNQGPPQ --- GPQGPPRPGNQOCPPPQ 233

Qy 101 FEEVKTPPEEGKQ 113
 Db 4 GGPQGPGRPNQQ 246

Matches 34; Conservative 15; Mismatches 36; Indels 22; Gaps 5;

Qy 15 RYVPEPMIGPMRPEQFSDEVEPATPEEGE -- -PATORQDPAAAQEGEDEGASAGQGPKP 71
 Db 47 RASEEEETGGP --- EDSEAVQPRTEPEPELEAATKGDEPDLQEMEPF ----- PEP 95

Qy 72 EAHSQEQQHPTQGCECEDGDQEMDPPNPEEVKTP - EGEKQSQC 116
 Db 96 EVPEPEPEPEPEPEPEB ----- PEPPEREDFQEGDESEGK 136

RESULT 5

S22373 proline-rich protein - mouse
 C;Species: Mus musculus (house mouse)
 C;Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 20-Aug-1999
 C;Accession: S22373
 R;Layfield, R.; Bannister, A.J.; Pierce, E.J.; McDonald, C.J.
 Eur. J. Biochem. 204, 591-597, 1992
 A;Title: cDNA clones for mouse parotid proline-rich proteins. mRNA regulation by
 A;Reference number: S22373; MUID:9174915; PMID:139347
 A;Accession: S22373
 A;Molecule type: mRNA
 A;Residues: 1-260 <LAY>
 A;Cross-references: EMBL:X63004; NID:953798; PIDN:CAA44733.1; PID:953799
 C;Superfamily: proline-rich protein
 C;Accession: EMBL:953799
 Query Match Score 105.5; DB 2; Length 260;
 Best Local Similarity 29.5%; Pred. No. 0.41;
 Matches 31; Conservative 15; Mismatches 50; Indels 9; Gaps 4;

Qy 11 PRPRRYVPEPMIGPMRPEQFSDEVEPATPEEGEPATQDPAAAQEGEDEGASAGQGPKP 70
 Db 144 PRFGNQQGPQGPQGPQGPQGPQGPQGPQGPQGPQGPQGPQ 196

Qy 71 --PEAHSQEQGHPTGTGCBEDGPQEMDPNNPEEVKTPPEEGKQ 113
 Db 197 GGHQPGNQGQGPQGPQGPQGPQGPQGPQGPQGPQGPQ 241

RESULT 6

A58938 surface protein rhotry ROP1 precursor - Toxoplasma gondii
 C;Species: Toxoplasma gondii
 C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C;Accession: A58938; A45644; S37697
 R;Boothroyd, J.C.
 Mol. Biochem. Parasitol. 50, 1-15, 1992
 A;Title: A Toxoplasma gondii rhotry protein associated with host cell penetratio
 A;Reference number: A45644; MUID:9178277; PMID:1542304
 A;Accession: A45644
 A;Molecule type: mRNA
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-396 <BOO>
 A;Cross-references: EMBL:M71274; NID:9897822; PIDN:AAA69859.1; PID:9897823
 A;Note: revision to sequence reported in A45644
 R;Ossorio, P.N.; Schwartzman, J.D.; Boothroyd, J.C.
 Mol. Biochem. Parasitol. 50, 1-15, 1992
 A;Title: A Toxoplasma gondii rhotry protein associated with host cell penetratio
 C;Keywords: surface antigen
 F;1-21/Domain: signal sequence #status predicted <SIG>
 F;22-396/Product: surface protein rhotry #status predicted <MAT>
 A;Residues: 'MACRQLCSYONLFFFLRDYCTDFDT', 1-352, 'FPQR', 358-364, 'R', 366, 'I', 393,
 A;Cross-references: EMBL:M71274; NID:9897822
 A;Note: sequence extracted from NCBI backbone (NCBIN:85178, NCBLIP:85179)
 C;Superfamily: surface protein rhotry
 C;Keywords: surface antigen
 F;1-21/Domain: signal sequence #status predicted <SIG>
 Query Match Score 104; DB 1; Length 396;
 Best Local Similarity 34.8%; Pred. No. 0.77;
 Matches 40; Conservative 8; Mismatches 41; Indels 26; Gaps 8;

Qy 9 YRPRPRRYVPEPMIGPMRPEQFSDEVEPATPEEGEPATQDPAAAQEGEDE 61
 Db 75 YEPLP-SFVEAP-WRGP --- DQV --- PARGEAALVTEETPAQOPAVAGSAEGEG 121

Qy 60 ---DEGASAGOGPKPEAHSQOEQGPQTGCEDGPQEMDPNNPEEVKTPPEGE 111
 Db 122 TSTTEASENSEDDTFHDALQEEDGLEVRP - PNAQELPPNVQELPPPTIEQ 175

RESULT 7

A27307 proline-rich phosphoprotein (gene PRH1; Db allele) - human
 C;Species: Homo sapiens (man)
 C;Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 29-Aug-1997
 C;Accession: A27307
 R;Azen, E.A.; Kim, H.S.; Goodman, P.; Flynn, S.; Maeda, N.
 Am. J. Hum. Genet. 41: 1035-1047, 1987
 A;Title: Alleles at the PRH1 locus coding for the human salivary-acidic proline-rich protein
 A;Reference number: A27307; MUID:88014305; PMID:3687941
 A;Status: nucleic acid sequence not shown
 A;Molecule type: DNA
 A;Residues: 1-171 <A2E>
 A;Cross-references: EMBL:K03203

C;Genetics:
 A;Gene: GDB:PRH1
 A;Cross-references: GDB:119515; OMIM:168730
 A;Map position: 12p13.2-12p13.2
 C;Superfamily: proline-rich protein
 C;Keywords: phosphoprotein

Query Match 16.0%; Score 103.5; DB 2; Length 171;
 Best Local Similarity 30.9%; Pred. No. 0.39;
 Matches 30; Conservative 9; Mismatches 41; Indels 17; Gaps 4;
 Query 29 EQPSDEVEPATPEEGPATQDPAAAQBGDEGASRQ-----QGPKPAAHSQSGQ 79
 Db 23 EQFLDEERGPPLGG---QOSQPSAGDGNDDGPQQGPPQQGQQQGP-PPPQGKEQG 77
 Query 80 HPQTSCCECSDG--PDGQEMDPNPEVKTPBEGEKQ 113
 Db 78 PPQQGQQQQSPPPPGKPPQCPPQQGHPPPQQGRPQ 114

RESULT 8
 B30843

Glutelin high molecular weight chain Ax2 precursor - wheat:

C;Species: Triticum aestivum (common wheat)
 C;Date: 19-May-1989 #sequence_revision 19-May-1989 #text_change 15-Nov-1996
 C;Accession: B30843
 R;Anderson, O.D.; Greene, F.C.
 Submitted to GenBank, January 1989
 A;Reference number: A94515
 A;Accession: B30843
 A;Molecule type: DNA
 A;Residues: 1-815 <AND>
 C;Superfamily: Glutelin
 C;Keywords: seed; storage protein
 F;1/1/Domain: signal sequence #status predicted <SIG>
 F;22-815/Product: Glutamine/Ax2 chain #status predicted <MAT>
 F;108-773/Region: glutamine/glycine/proline-rich

Query Match 15.9%; Score 102.5; DB 2; Length 815;
 Best Local Similarity 23.7%; Pred. No. 1.9;
 Matches 31; Conservative 27; Mismatches 54; Indels 19; Gaps 5;
 Query 2 SWGRSTYRPERRYEPPEMIG-----PMRPEQFSDEVEPATPEEGPATQDPA 54
 Db 362 SGQQQGPQRGPQGYTSTSQQLGQGPQRYPTSPQQPQEQPQQGPQQPQ 421
 Query 55 AQEGE----DEGASAGQGPKEAHSSQEGHPTG--HPOTGCCEDGPDGQEMDPNPPE--EV 104
 Db 422 GQGQGQORGQGQGPQGPQGPQGPQGPQGPQGPQGPQGPQGPQGPQGPQGPQ 479
 Query 105 KTPPEGEKQSQ 115
 Db 480 QQPACGQQGPQ 490

RESULT 9
 JN0889
 glutelin, high-molecular-weight Ax2* chain precursor - wheat

C;Species: Triticum aestivum (common wheat)
 C;Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 15-Nov-1996
 C;Accession: JN0889
 R;Anderson, O.D.; Greene, F.C.
 Theor Appl. Genet. 77, 689-700, 1989
 A;Title: The characterization and comparative analysis of high-molecular-weight
 A;Reference number: JN0889
 A;Accession: JN0889
 A;Molecule type: DNA
 A;Residues: 1-815 <AND>
 A;Note: the authors translated the codon CTA for residue 11 as Val. CAT for residue
 C;Comment: The main wheat storage proteins are divided into two groups. The glut
 a families.
 C;Genetics:
 A;Gene: Ax2*

C;Superfamily: Glutelin
 F;1-21/Domain: signal sequence #status predicted <SIG>
 F;22-815/Product: Glutelin, high-molecular-weight Ax2* chain #status predicted <
 Query Match 15.9%; Score 102.5; DB 2; Length 815;
 Best Local Similarity 23.7%; Pred. No. 1.9; Gaps 5;
 Matches 31; Conservative 27; Mismatches 54; Indels 19; Gaps 5;
 Query 2 SWGRSTYRPERRYEPPEMIG-----PMRPEQFSDEVEPATPEEGPATQDPA 54
 Db 362 SGQQQGPQRGPQGYTSTSQQLGQGPQRYPTSPQQPQEQPQQGPQQPQ 421
 Query 55 AQEGE----DEGASAGQGPKEAHSSQEGHPTG--HPOTGCCEDGPDGQEMDPNPPE--EV 104
 Db 422 GQGQGQORGQGQGPQGPQGPQGPQGPQGPQGPQGPQGPQGPQGPQGPQ 479
 Query 105 KTPPEGEKQSQ 115
 Db 480 QQPACGQQGPQ 490

RESULT 11
 S02262
 glutelin high molecular weight chain Dx5 - wheat

C;Species: Triticum aestivum (common wheat)
 C;Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 03-Feb-1994
 C;Accession: S02262
 R;Anderson, O.D.; Greene, F.C.; Yip, R.E.; Halford, N.G.; Shewry, P.R.; Malpica-Romero,
 Nucleic Acids Res 17, 461-462, 1989
 A;Title: Nucleotide sequences of the two high-molecular-weight glutenin genes from the D
 A;Reference number: S02262; MUID:89098419; PMID:2563152
 A;Accession: S02262
 A;Molecule type: DNA
 A;Residues: 1-848 <AND>
 A;Cross-references: EMBL:X12928
 C;Genetics:
 A;Gene: Glu-D1-b
 C;Superfamily: glutenin

Query Match Score 15.9%; DB 2; Length 648;
 Best Local Similarity 23.4%; Pred. No. 2;
 Matches 33; Conservative 24; Mismatches 53; Indels 31; Gaps 6;

Qy 4 RGRSTYRP-----RPRRTVEP-----MIGPMRPEQFS---DEVEPATPEEGE 44
 Db 491 QGPGYRPFPSQOSGOGPQGQMQGPQGQPGYRPFSPQPGQGPYDTSQPGQGQ 550

Qy 45 PATQRDPAAQEGE-----DEGASAGQGPKE--AHSQEQGHPORTCECDBPDGQ 94
 Db 551 QPGQLOOAQGQDQQAQGQOQPAQVQGQRAQGQGQPGQGQQLGQGQGQ 610

Qy 95 EMDDPPNDEEVKTPPEEGEKSQ 115
 Db 611 Q---PGQGQQGOQPAQGQGQ 629

RESULT 12
 PC4014 calcium binding 140k protein - mouse (fragment)
 C;Species: Mus musculus (house mouse)
 C;Date: 21-Jul-1995 #sequence_revision 31-Mar-1997 #text_change 05-Nov-1999
 C;Accession: PC014
 R;Naved, A.F.; Ozawa, M.; Yu, S.; Miyuchi, T.; Muramatsu, H.; Muramatsu, T.
 Cell Struct. Funct. 20, 133-141, 1995
 A;Title: CBP-140, a novel endoplasmic reticulum resident Ca2+-binding protein with a car
 A;Reference number: PC4014; MUID:95368840; PMID:1641295
 A;Accession: PC4014
 A;Molecule type: mRNA
 A;Residues: 1-652 <NAV>
 A;Cross-references: GB:S878797; NID:91041986; PID:AB250511; PID:91041987
 C;Comment: The sequence 1-80 has high homology to that of 70k heat shock protein.
 C;Keywords: calcium binding endoplasmic reticulum; heat shock; stress-induced protein
 F;649-652/Region: endoplasmic reticulum retention signal #statast atypical

Query Match Score 15.8%; DB 2; Length 652;
 Best Local Similarity 34.1%; Pred. No. 1.7;
 Matches 31; Conservative 8; Mismatches 36; Indels 16; Gaps 4;

Qy 25 PMRPEQFSDEVEPATPEEGEPATQRQDPAAQEGE--DEGASAGQGPKEAHQSQEQQHPPQ 82
 Db 270 PAEQGELKEAEPAAETSSQPPSEPKGDAAREGEKPKDKES--GDKPAQPNENKG-- 323

Qy 83 TGCECEGQPDGQEMDPNPNEVKTPEEGEKKQ 113
 Db 324 -----QAGPEGA---APAPEEDKLKPARKQ 346

RESULT 13
 T30282 calcium-binding protein - sea urchin (Strongylocentrotus purpuratus)
 C;Species: Strongylocentrotus purpuratus (purple urchin)
 C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 11-May-2000
 C;Accession: T30282
 R;Soleysik-Espanol, M.; Klinzing, D.C.; Pfarr, K.; Burke, R.D.; Ernst, S.G.
 Dev. Biol. 165, 73-85, 1994
 A;Title: Endo16, a large multidomain protein found on the surface and ECM of endodermal
 A;Reference number: Z20805; MUID:94374583; PMID:8088452

A;Accession: T30282
 A;Status: preliminary; translated from GB/EMBL/DDJB
 A;Molecule type: mRNA
 A;Cross-references: <SOL>
 A;Cross-references: EMBL:L34680; NID:9511893; PID:9511894; PID:AAA30047.1
 C;Genetics:
 A;Note: endo16

Query Match Score 15.8%; DB 2; Length 1560;
 Best Local Similarity 32.0%; Pred. No. 3.8;
 Matches 32; Conservative 13; MisMatches 35; Indels 20; Gaps 5;

Qy 34 EVPATPEEGEPATQRQDPAAQEGEDEGASAGQGPKEAHQSQEQQHPPQGCRCB---- 88
 Db 1362 EVKTSNQEESEVQEQSDQFQEMEEGQG-SSEGEPEPEPEEGQ-PQEQQESEGEENE 1419

Qy 89 -----DGP-----DGQMDPPNP-EVTPEEGEKSQ 115
 Db 1420 SEEVEVEPQGVMESEGOSESEDEPEGEPESEGEQE 1459

RESULT 14
 S11560 proline-rich protein MP4 - mouse
 C;Species: Mus musculus (house mouse)
 C;Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 20-Aug-1999
 C;Accession: S11560; S22570
 R;Roberts, S.G.E.; Layfield, R.; Bannister, A.J.; McDonald, C.J.
 Eur. J. Biochem. 202, 969-974, 1991
 A;Title: Gene sequence of mouse B-type proline-rich protein MP4. Transcriptional
 A;Reference number: S11560; MUID:92111548; PMID:1765104
 A;Accession: S11560
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-1300 <ROB>
 A;Cross-references: GB:X58438; NID:953181; PID:CAA41344.1; PID:953182
 R;Roberts, S.G.E.; Layfield, R.; McDonald, C.J.
 Nucleic Acids Res. 19, 5205-5211, 1991
 A;Title: The mouse proline-rich protein MP6 promoter binds isoprenaline-inducible
 A;Reference number: S22570
 A;Accession: S22570
 A;Molecule type: DNA
 A;Residues: 1-14 <RO2>
 A;Cross-references: EMBL:X61126
 C;Superfamily: proline-rich protein

Query Match Score 15.7%; DB 2; Length 300;
 Best Local Similarity 28.8%; Pred. No. 0.91;
 Matches 38; Conservative 10; MisMatches 55; Indels 29; Gaps 7;

Qy 4 RGRSTYR-PRRRTYVEPPPEMIG-MRPEQFSDEVEPATPEEGEPATQRQDPAAQEGEDE 61
 Db 121 QGSQQRPPQPGNGQGPQGPQGPQGPQGPQGPQGPQGPQGPQGPQGPQGPQ 178

Qy 62 GA-----SAGQGPKEAHQSQ-----EGHPTGCECEDGPD----QEMDPPNP 101
 Db 179 GGPPIPREGNQQGPQGPQGPQGPQGPQGPQGPQGPQGPQGPQGPQGPQGPQ 233

Qy 102 EYKTPPEEGEKO 113
 Db 234 GPQGPPTRNQOQ 245

RESULT 15
 B97586 hypothetical protein AGR_C_3445 [imported] - Agrobacterium tumefaciens (strain C;Species: Agrobacterium_tumefaciens
 C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
 C;Accession: B97586
 R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.;
 A.; Liu, F.; Wollam, C.; Scott, C.; Doughty, D.; Allinger, M.; Lappas, C.; Mar
 Science 294, 2323-2328, 2001
 A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobact-

A;Reference number: A97359; PMID:11743194
A;Accession: B97566
A;Status: Preliminary
A;Molecule type: DNA
A;Residues: 1-754 <KUR>
A;Cross-references: GB:AE007869; PIDN:AAK87643.1; PID:g15156995; GSDB:GN00169
C;Genetics:
A;Gene: AGR C 3445
A;Map position: circular chromosome

Query Match 15 6%; Score 101; DB 2; Length 754;
Best Local Similarity 21.7%; Pred. No. 2.3;
Matches 36; Conservative 15; Mismatches 47; Indels 32; Gaps 5;
Qy 10 RPRPRRYVTPPEPMIGPMRQEFSDEVEPATPPEEGEBATQDPAAAQEGEDGASAGQ-- 67
Db 182 RPRKPKQAEQPAGEQRERPRKEKEPK--EPAAEQQ-PAARPENAEPAKPREPA 236
Qy 68 -GPKPBEAHSQEOGHPQTGCCECDGPDQMD----- PPNPPEVK 105
Db 237 PGKKPQV--EKAPEQKAEPAEKVPEKKPAAPPAKEAPVPTAATPAPPPAPPAPQP 293
Qy 106 TPEEGEKQSQ 115
Db 294 NPAPGROPSB 303

Search completed: April 7, 2003, 18:14:23
Job time : 46 secs



Result No.	Score	Query Match	Length	DB ID	Description
1	646	100.0	116	1	G352_HUMAN
2	637	98.6	116	1	G358_HUMAN
3	625.5	96.8	117	1	G357_HUMAN
4	616.5	95.4	117	1	G354_HUMAN
5	612.5	94.8	117	1	G355_HUMAN
6	608.5	94.2	117	1	G356_HUMAN
7	599	92.6	138	1	G351_HUMAN
8	589.5	91.3	118	1	G353_HUMAN
9	266	41.2	146	1	G351_HUMAN
10	229.5	35.5	111	1	G353_HUMAN
11	120	18.6	102	1	G351_HUMAN
12	110	17.0	221	1	N03_RAT
13	100.5	15.6	605	1	GLCA_SOYEN
14	99.5	15.4	548	1	NFL_FIG
15	99	15.3	839	1	GLUTS_WHEAT
16	98	15.2	838	1	GLUT4_WHEAT
17	97.5	15.1	166	1	PRIC_HUMAN
18	97.5	15.1	220	1	N03_MOUSE
19	97.5	15.1	235	1	BAR6_CHITE
20	97.5	15.1	541	1	NFL_RAT
21	97	15.0	261	1	PRC2_MOUSE
22	96.5	14.9	229	1	VE4 HPV08
23	96.5	14.9	617	1	ABP1_SACEX
24	96	14.9	542	1	NFL_MOUSE
25	96	14.9	554	1	NFL_BOVIN
26	95.5	14.8	543	1	NFL_HUMAN
27	95.5	14.8	634	1	SCL_RAT
28	94.5	14.6	245	1	VE4 HPV05
29	94.5	14.6	247	1	PRB4_HUMAN
30	94	14.6	234	1	PREM_HUMAN
31	93.5	14.5	276	1	PRL_HUMAN
32	93.5	14.5	580	1	SYN3_HUMAN
33	93	14.4	574	1	PRP_HUMAN

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 7, 2003, 18:13:10 ; Search time 12 seconds
(without alignments)

400.938 Million cell updates/sec

Title: US-09-782-745-27

Perfect score: 646

Sequence: 1 NSWRGRSTYRPRPRRYVEPP DPPNPEEVKTPPEEGEKQSQC 116

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112392 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing First 45 summaries

Database : SwissProt_40.4

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	646	100.0	116	1	G352_HUMAN
2	637	98.6	116	1	G358_HUMAN
3	625.5	96.8	117	1	G357_HUMAN
4	616.5	95.4	117	1	G354_HUMAN
5	612.5	94.8	117	1	G355_HUMAN
6	608.5	94.2	117	1	G356_HUMAN
7	599	92.6	138	1	G351_HUMAN
8	589.5	91.3	118	1	G353_HUMAN
9	266	41.2	146	1	G351_HUMAN
10	229.5	35.5	111	1	G353_HUMAN
11	120	18.6	102	1	G351_HUMAN
12	110	17.0	221	1	N03_RAT
13	100.5	15.6	605	1	GLCA_SOYEN
14	99.5	15.4	548	1	NFL_FIG
15	99	15.3	839	1	GLUTS_WHEAT
16	98	15.2	838	1	GLUT4_WHEAT
17	97.5	15.1	166	1	PRIC_HUMAN
18	97.5	15.1	220	1	N03_MOUSE
19	97.5	15.1	235	1	BAR6_CHITE
20	97.5	15.1	541	1	NFL_RAT
21	97	15.0	261	1	PRC2_MOUSE
22	96.5	14.9	229	1	VE4 HPV08
23	96.5	14.9	617	1	ABP1_SACEX
24	96	14.9	542	1	NFL_MOUSE
25	96	14.9	554	1	NFL_BOVIN
26	95.5	14.8	543	1	NFL_HUMAN
27	95.5	14.8	634	1	SCL_RAT
28	94.5	14.6	245	1	VE4 HPV05
29	94.5	14.6	247	1	PRB4_HUMAN
30	94	14.6	234	1	PREM_HUMAN
31	93.5	14.5	276	1	PRL_HUMAN
32	93.5	14.5	580	1	SYN3_HUMAN
33	93	14.4	574	1	PRP_HUMAN

ALIGNMENTS

RESULT 1									
ID	G352_HUMAN	STANDARD;	PRT;	116	AA.				
AC	Q13066;								
DT	15-JUL-1998	(Rel. 36, Created)							
DT	15-JUL-1998	(Rel. 36, Last sequence update)							
DE	GAGE2_protein (G antigen 2).								
GN	GAGE2.								
OS	Homo sapiens (Human).								
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Buteridae; Primates; Catarrhini; Hominidae; Homo.								
NCBI_TaxID	9606;								
RA	van den Bynde B., Peeters O., de Backer O., Gaugler B., Lucas S., Boon T.,								
RA	"A new family of genes coding for an antigen recognized by autologous cytolytic T lymphocytes on a human melanoma.",								
CC	-I- FUNCTION: ANTIGEN, RECOGNIZED ON MELANOMA BY AUTOLOGOUS CYTOLYTIC T LYMPHOCYTES								
CC	-I- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF TUMOR TISSUES BUT NOT IN NORMAL TISSUES, EXCEPT TESTIS.								
CC	-I- SIMILARITY: BELONGS TO THE GAGE FAMILY.								
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).								
CC	EMBL; U11443; AAA82745.1; -.								
DR	GeneID: HGENC:4093; GAGE2.								
DR	MINI; 604244;								
KW	Multigene family; Antigen.								
SQ	SEQUENCE 116 AA; 12786 MW; DD305DPCA29AF19A CRC64;								
QY	1 MSWGRSTYRPRYYVEPPMIGPQRPEFSEVATPTEEEGPATQRDPAAQGED 60								
Db	1 MSWGRSTYRPRYYVEPPMIGPQRPEFSEVATPTEEEGPATQRDPAAQGED 60								
QY	61 EGASAGQPKPEAHISOEQHQTGCCEBDGPQOEMDPNPETVTKPEGEKQSQC 116								
Db	61 EGASAGQPKPEAHISOEQHQTGCCEBDGPQOEMDPNPETVTKPEGEKQSQC 116								

RESULT 2

-1- SIMILARITY: BELONGS TO THE GAGE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration - between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions. This statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>) .

CC DR EMBL: U19145; AAA82747.1; -.

CC DR GeneID: HGNC:101; GAGE4.

CC DR MIM: 604246; -.

KW Multigene family.

SQ SEQUENCE 117 AA; 12956 MW; 365ED71B2F9DC7AD CRC64;

RESULT 5

Query Match Best Local Similarity	95.4%	Score 616.5;	DB 1;	Length 117;
Mismatches	113;	Pred. No.	3.6e-36;	
Matches	Conservative	1;	Mismatches	2;
			Indels	1;
			Gaps	1;

Qy 1 MSWRGRSTYR-PRPRRYVPEPMIGPMRPEQSDEVEPATPGEPEPATQRDPAAAQEAE 59

Db 1 MSWRGRSTYR-PRPRRYVPEPMIGPMRPEQSDEVEPATPGEPEPATQRDPAAAQEAE 60

Qy 60 DEGASAGQGPKPKEAHSQEQGHFQTCEDGPDGQEMDPNPNEEVKTPGEPEKOSQC 116

Db 61 DEGASAGQGPKPKEADSQEQGHFQTCEDGPDGQEMDPNPNEEVKTPGEPEKOSQC 117

RESULT 6

Query Match Best Local Similarity	94.8%	Score 612.5;	DB 1;	Length 117;
Mismatches	112;	Pred. No.	6.8e-36;	
Matches	Conservative	2;	Mismatches	2;
			Indels	1;
			Gaps	1;

Qy 1 MSWRGRSTYR-PRPRRYVPEPMIGPMRPEQSDEVEPATPGEPEPATQRDPAAAQEAE 59

Db 1 MSWRGRSTYR-PRPRRYVPEPMIGPMRPEQSDEVEPATPGEPEPATQRDPAAAQEAE 60

Qy 60 DEGASAGQGPKPKEAHSQEQGHFQTCEDGPDGQEMDPNPNEEVKTPGEPEKOSQC 116

Db 61 DEGASAGQGPKPKEADSQEQGHFQTCEDGPDGQEMDPNPNEEVKTPGEPEKOSQC 117

RESULT 7

Query Match Best Local Similarity	94.2%	Score 608.5;	DB 1;	Length 117;
Mismatches	111;	Pred. No.	1.3e-35;	
Matches	Conservative	3;	Mismatches	2;
			Indels	1;
			Gaps	1;

Qy 1 MSWRGRSTYR-PRPRRYVPEPMIGPMRPEQSDEVEPATPGEPEPATQRDPAAAQEAE 59

Db 1 MSWRGRSTYR-PRPRRYVPEPMIGPMRPEQSDEVEPATPGEPEPATQRDPAAAQEAE 60

Qy 60 DEGASAGQGPKPKEADSQEQGHFQTCEDGPDGQEMDPNPNEEVKTPGEPEKOSQC 116

Db 61 DEGASAGQGPKPKEADSQEQGHFQTCEDGPDGQEMDPNPNEEVKTPGEPEKOSQC 117

DT 15-JUL-1998 (Rel. 36; Last sequence update)
 DT 30-MAY-2000 (Rel. 39; Last annotation update)
 DE GAGE1.
 GN Homo sapiens (Human).
 OS Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TAXID=9606;
 RN [1] SEQUENCE FROM N.A.
 RC TISSUE-Melanoma;
 RX MEDLINE=95378788; PubMed=7544395;
 RA van den Eynde B., Peeters O., de Backer O., Gaugler B., Lucas S., Boon T.;
 RA RT A new family of genes coding for an antigen recognized by autologous cytolytic T lymphocytes on a human melanoma.;
 RL J. Exp. Med. 182:689-698(1995).
 CC -!- FUNCTION: ANTIGEN, RECOGNIZED ON MELANOMA BY AUTOLOGOUS CYTOLYTIC T LYMPHOCYTES. COMPLETELY SILENT IN NORMAL ADULT TISSUES, EXCEPT TESTIS.
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF TUMOR TISSUES BUT NOT IN NORMAL TISSUES, EXCEPT TESTIS.
 CC -!- SIMILARITY: BELONGS TO THE GAGE FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR U19142; AAA82744.1; .
 DR Genew; HGNC:4098; GAGE1.
 DR MM: 604243; .
 KW Multigene Family; Antigen.
 SEQUENCE 138 AA; 15418 MW; 3788F3909EC4B3B2 CRC64;

Query Match Score 92.6%; DB 1; Length 138;
 Best Local Similarity 99.1%; Pred. No. 7.6e-35;
 Matches 108; Conservative 0; Missmatches 1; Indels 0; Gaps 0;

Qy 1 MSWRGRSTYRPRRYVEPEPMIGPMRPPQFSDEVPAPPEEGEPATORQDPAAQEGED 60
 Db 1 MSWRGRSTYRPRRYVEPEPMIGPMRPPQFSDEVPAPPEEGEPATORQDPAAQEGED 60

Qy 61 EGASAGGCPKPEAHSSDQGHPTGCECIGPDGOEMDPNPEEVKTPPE 109
 Db 61 EGASAGGCPKPEADSDQGHPTGCECIGPDQEMDPNPEEVKTPPE 109

RESULT 8
 GAGE3_HUMAN ID GGE3_HUMAN STANDARD; PRT; 118 AA.
 AC Q13067;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE GAGE-3 protein (G antigen 3).
 GN Homo sapiens (Human).
 OS Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TAXID=9606;
 RN [1] SEQUENCE FROM N.A.
 RC TISSUE-Melanoma;
 RX MEDLINE=95378788; PubMed=7544395;
 RA van den Eynde B., Peeters O., de Backer O., Gaugler B., Lucas S., Boon T.;
 RA RT A new family of genes coding for an antigen recognized by autologous cytolytic T lymphocytes on a human melanoma.;
 RL J. Exp. Med. 182:689-698(1995).

CC -!- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF TUMOR TISSUES BUT NOT IN NORMAL TISSUES, EXCEPT TESTIS.
 CC -!- SIMILARITY: BELONGS TO THE GAGE FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR U19144; AAA82746.1; .
 DR Genew; HGNC:4100; GAGE3.
 DR MM: 604245; .
 KW Multigene family;
 SEQUENCE 118 AA; 12937 MW; D97EBBB19E735103 CRC64;

Query Match Score 91.3%; DB 1; Length 118;
 Best Local Similarity 94.7%; Pred. No. 2.5e-34;
 Matches 108; Conservative 3; Missmatches 2; Indels 1; Gaps 1;

Qy 4 RGRSTYR-PRPRYYVEPEPMIGPMRPPQFSDEVPAPPEEGEPATORQDPAAQEGEDG 62
 Db 5 RGKSTYYWPRRTYQPPETVIGPMRPPQFSDEVPAPPEEGEPATORQDPAAQEGEDG 64

Qy 63 ASAGQGPKPKEAHSSDQGHPTGCECIGPDQEMDPNPEEVKTPPEEGEKOSOC 116
 Db 65 ASAGQGPKPKEADSDQGHPTGCECIGPDQEMDPNPEEVKTPPEEGEKQSSC 118

RESULT 9
 GGB1_HUMAN STANDARD; PRT; 146 AA.
 ID GGB1_HUMAN STANDARD;
 AC O75459; O9BSS7;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)

DB G antigen family B 1 protein (Prostate-associated gene protein 1)
 DE (PAGE-1) (GAGE-9) (AL5).
 GN GAGEB1 OR PAGE1 OR GAGE9.
 OS Homo sapiens (Human).
 OC Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Buteleostomi;
 OC Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=95606;
 RN SEQUENCE FROM N.A.
 RP MEDLINE=9651357;
 RX RA Chen M.E., Lin S.-H., Chung L.W.K., Sikes R.A.;
 RA "Isolation and characterization of PAGE-1 and GAGE-7: new genes expressed in the LNCaP prostate cancer progression model that share homology with melanoma associated antigens.",
 RT J. Biol. Chem. 273:17618-17625(1998).
 RT J. Biol. Chem. 273:17618-17625(1998).
 RN SEQUENCE FROM N.A.
 RP RC TISSUE=Skin;
 RA Strausberg R.; Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 RL -!- FUNCTION: UNKNOWN.
 CC -!- TISSUE SPECIFICITY: ISOLATED FROM PROSTATE CANCER CELL LINES;
 CC EXPRESSION ASSOCIATED WITH PROGRESSION TO ANDROGEN INSENSITIVE PHENOTYPE. EXPRESSED IN NORMAL TESTIS AND AT LOWER LEVEL IN NORMAL PLACENTA.
 CC -!- SIMILARITY: BELONGS TO THE GAGE FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR	EMBL; AF058989; AAC25990; 1;	Qy	1 MSWGRSRTYPRPRRYVEPPMIGPMRPEQFSIDEVEPA -TPPEGEPATORDQDPAAGQG 58
DR	EMBL; BC004861; AAH04861; -;	Db	1 MSWGRSRTYPRPRRSLOQQPLLGAM-----LEPTDEEPEKEPKPDKR 52
DR	Genew; HGNC:4107; GAGEB1.		
DR	MIM: 300288;		
KW	Multigene family.		
FT	CONFLICT 75 75 MW; P -> L (IN REF. 2).		
FT	SEQUENCE 146 AA; 16134 MW; 91ABALE3B498DCA2 CRC64;		
Query Match	41.2%; Score 266; DB 1; Length 146;		
Best Local Similarity 42.0%; Pred. No. 3.2e-12;			
Matches 66; Conservative 7; Mismatches 30; Indels 54; Gaps 3;			
GGCD3_HUMAN STANDARD; PRT; 102 AA.			
AC Q96GT9; DT 15-JUN-2002 (Rel. 41, Created)			
DB 1 MGFLRLRIRRMLYY-----SSEESDE---QPDEVESPQSQDSTAERED 48			
Qy 61 EGASAGO-----GPKPEAHSSQ 78			
DB 49 EGASAAQGPEADSQLELVQRTGCPGDGPDTKRVCLNEEQMKLPAEGPEPEADSQ 108			
Qy 79 GHQTGCCEGDGPQCEMDPNPPEEVTKTPPEGEKOSQ 115			
DB 109 VHPKTGCERGDGPDVQELGLPNPPEEVTKTPPEEDGQSQ 145			
RESULT 10			
GGCD3_HUMAN STANDARD; PRT; 111 AA.			
AC Q96GT9; DT 15-JUN-2002 (Rel. 41, Created)			
DB 15-JUN-2002 (Rel. 41, Last sequence update)			
DT 15-JUN-2002 (Rel. 41, Last annotation update)			
DS G antigen family D 3 protein (XAGE-2).			
OS Homo sapiens (Human).			
GN Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;			
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX NCBI_TaxID=9606;			
RN [1]			
SEQUENCE FROM N.A.			
RX MEDLINE=38393718; PubMed=9724777;			
RA Brinkmann U., Vasmatis G., Lee B., Yerushalmi N., Essand M.,			
RA Pascan I.;			
RA "PAGE-1, an X chromosome-linked GAGE-like gene that is expressed in normal and neoplastic prostate, testis, and uterus.";			
RT Proc. Natl. Acad. Sci. U.S.A. 95:10757-10762(1998).			
RA [2]			
SEQUENCE FROM N.A.			
RA Strom T.M., Nyakatira G., Hellebrand H., Drescher B., Rosenthal A.,			
RA Meindl A.;			
RA Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.			
RA [3]			
SEQUENCE FROM N.A.			
RC TISSUE=Placenta;			
RA Strausberg R.;			
RL Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.			
-I- FUNCTION: UNKNOWN.			
CC -I TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN NORMAL MALE AND			
CC FEMALE REPRODUCTIVE TISSUES, PROSTATE, TESTIS, FALLOPIAN TUBE,			
CC UTERUS, AND PLACENTA, AS WELL AS IN PROSTATE, CANCER, TESTICULAR			
CC CANCER, AND UTERINE CANCER.			
CC -I SIMILARITY: BELONGS TO THE GAGE FAMILY.			
CC			
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC DR EMBL; AF75258; AAP88037.1; -.			
CC DR EMBL; AJ05894; CAA06751.1; -.			
CC DR EMBL; AF338380; AAF2541.1; -.			
CC DR EMBL; BC010897; AAH10897.1; -.			
CC DR Genew; HGNC:4108; GAGEC1.			
CC DR MIM: 300387; -.			
CC DR KW Multigene family.			
SQ SEQUENCE 102 AA; 11153 MW; CEB5D07AFBF73301B CRC64;			
Query Match 18.6%; Score 120; DB 1; Length 102;			
Best Local Similarity 34.5%; Pred. No. 0.02;			
Matches 39; Conservative 10; Mismatches 52; Indels 12; Gaps 3;			
Qy 1 MSWGRSRTYPRPRRYVEPPMIGPMRPEQFSIDEVEPA -TPPEGEPATORDQDPAAGQG 60			
Db 1 MSARVRSRSRGDGQ-EAPDVVA-----FVAPGESQQEEPTDNODIEPQG--ER 48			

OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC	euRosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX	
RN	SEQUENCE FROM N.A.
RP	TISSUE-Cotyledon;
RC	MEDLINE=91358860; PubMed=2103438;
RX	Sebastiani F.L., Farre L.B., Schuler M.A., Beachy R.N.;
RA	"Complete sequence of a cDNA of alpha subunit of soybean beta-a conglycinin";
RT	Plant Mol Biol. 15:197-201(1990).
RL	-!- FUNCTION: SEED STORAGE PROTEIN. ACCUMULATES DURING SEED DEVELOPMENT AND IS HYDROLYZED AFTER GERMINATION TO PROVIDE A CARBON AND NITROGEN SOURCE FOR THE DEVELOPING SEEDLING.
CC	-!- SUBUNIT: THE ALPHA -, ALPHA -, AND BETA-SUBUNITS ASSOCIATE IN VARIOUS COMBINATIONS TO FORM TRIMERIC PROTEINS.
CC	-!- SUBCELLULAR LOCATION: EMBRYO AXIS, AND COTYLEDONY MEMBRANE-BOUND VACUOLAR PROTEIN BODIES.
CC	-!- SIMILARITY: BELONGS TO THE 7S SEED STORAGE PROTEIN FAMILY.
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on use by non-profit institutions as long as its content is in no modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcements or send an email to license@isb-sib.ch).
CC	
DR	EMBL; X17698; CAA5691.1; -.
DR	PIR; S14681; FWSYBA..
HSSP	P50477; 1CAU
DR	InterPro; IPR001113; Seedstore_7s.
DR	PFam; PF00546; Seedstore_7s_1.
DR	PFam; PF02808; Seedstore_7s_C_1.
KW	Seed storage protein, Signal, Glycoprotein, Multigene family.
FT	SIGNAL 1 22
FT	PROPEP 23 62
FT	CHAIN 63 605 BETA-CONGLYCININ, ALPHA CHAIN.
FT	CARBHYD 261 261 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBHYD 517 517 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE 605 AA; 70293 MW; CBEBA30506BBCBC57 CRC64;
Query	Match Score 100.5; Length 605;
Best Local Similarity 28.1%; Pred. No. 2.1;	
Matches 34; Conservative 12; Mismatches 40; Indels 35; Gaps 62	
Qy	10 RPRPRRYVEPPPEMIGPMRPFQFSDEVEPATFEEGEPATQRQD-----PAAAQEGEDEG
Db	75 RPRPR-----PQHPER-----EPQQQPEKEEDEDQPRPDTIPPRQPRQEERHEQ 11.9
Qy	63 ASAGQGPPEAHSSQFGHPOTGCBEDGPDGQBM-----PPNPEEVKTPBEGBKQS 11.4
Db	120 REQEQWPR----KPEKRQGKGSEEEDEDDEQDERQFPEPRPHQKEPRNEEEDEE 17.4
Qy	115 Q 115
Db	175 Q 175
RESULT 14	
NFL_PIG	STANDARD;
ID	NP_02547;
AC	NP_02547;
DT	21-JUL-1986 (Rel. 01, Created)
DT	01-OCT-1996 (Rel. 34, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Neurofilament triplet L protein (68 kDa neurofilament protein) (NF-L).
GN	NEFL.
OS	Sus scrofa (pig).
OC	Eukarya; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

OX NCBI_TaxID=9823;
 RN [1] sequence;
 RP
 RX MEDLINE=85114583; PubMed=3920075;
 RA Geisler N., Plessmann U.; Weber K.;
 RT "The complete amino acid sequence of the major mammalian neurofilament protein (NF-L).";
 RL FEBS Lett. 182:475-478 (1985).
 RN [2] sequence of 1-82 AND 278-548.
 RA Geisler N., Kaufmann B., Fischer S., Plessmann U., Weber K.;
 RT "Neurofilament architecture combines structural principles of intermediate filaments with carboxy-terminal extensions increasing in size between triplet proteins.";
 RL EMBO J. 2:1955-1962 (1983).
 CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M, H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
 CC -1- DOMAIN: THE EXTRA MASS AND HIGH CHARGE DENSITY THAT DISTINGUISH THE NEUROFILAMENT PROTEINS FROM ALL OTHER INTERMEDIATE FILAMENT PROTEINS ARE DUE TO THE TAILPIECE EXTENSIONS. THIS REGION MAY FORM A CHARGED SCAFFOLDING STRUCTURE SUITABLE FOR INTERACTION WITH OTHER NEURONAL COMPONENTS OR IONS.
 CC -1- MISCELLANEOUS: NF-L IS THE MOST ABUNDANT OF THE THREE NEUROFILAMENT PROTEINS AND, AS THE OTHER NONEPITHELIAL INTERMEDIATE FILAMENT PROTEINS, IT CAN FORM HOMOPOLYMERIC 1.0-NM FILAMENTS.
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 DR PIR; A02963; QFPGL.
 DR InterPro; IPR01664; IP.
 CC Pfam: PF00038; filament; 1.
 DR PROSITE; PS00226; IF; 1.
 KW Intermediate filament; Coiled coil; Neuron; Glycoprotein.
 FT DOMAIN 1 91 HEAD
 FT DOMAIN 92 91 ROD.
 FT DOMAIN 396 548 TAIL.
 FT DOMAIN 92 123 COIL 1A.
 FT DOMAIN 124 136 LINKER 1.
 FT DOMAIN 137 232 COIL 1B.
 FT DOMAIN 233 251 LINKER 12.
 FT DOMAIN 252 270 COIL 2A.
 FT DOMAIN 271 279 LINKER 2.
 FT DOMAIN 395 442 TAIL, SUBDOMAIN A.
 FT DOMAIN 443 548 TAIL, SUBDOMAIN B (ACIDIC).
 FT CARBHYD 20 20 O-LINKED (GLCNAC) (BY SIMILARITY).
 FT CARBHYD 26 26 O-LINKED (GLCNAC) (BY SIMILARITY).
 FT SITE 380 390 EPTOPE (RECOGNIZED BY IF-SPECIFIC MONOClonal ANTIBODY).
 FT UNSURE 322 322 OR K.
 SQ SEQUENCE 548 AA; 61940 MW; 83044813637A6739 CRC64;
 Query Match 15.4%; Score 99.5%; DB 1; Length 548;
 Best Local Similarity 26.2%; Pred. No. 2.2; Mismatches 54; Indels 25; Gaps 4;
 Matches 34; Conservative 17; MisMatches 54; Indels 25; Gaps 4;

QY 5 GRSTY-----RPRPRY-----VPPENTGPMRBPQFSDEVEPATPEEGE 44
 DB 419 GRSAYGGLQTSYLMSTRFSPSYTSHVQQEEVETEAAKEEAE 477

QY 45 PATQRDPAAQEGEDGEGASAGCQGPKEAHQSFGHPTGCECDGPDQEMDPENPEEV 104
 DB 478 EGKEKEEAEEAEAEAEAGAQSEEEAAKEESEAKEEBEGG---EGEQBETKEAEEEK 533

QY 105 KTPPEGEKQS 114
 DB 534 KDEGAGEEQQA 543

RESULT 15
 GITS_WHEAT STANDARD; PRT; 839 AA.
 ID GLUT5 WHEAT
 AC P10388;
 DT 01-MAR-1989 (Rel. 10, Created)

01-NOV-1997 (Rel. 35, Last sequence update)
 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Glutinin, high molecular weight subunit Dx5 precursor.
 GN Glu-11D OR Glu-D1-B.
 OS Triticum aestivum (Wheat).
 OC Spermatophyte; Viriliplantae; Streptophytina; Embryophytina; Tracheophytina; Magnoliophytina; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
 OC NCBI_TaxID=4365;
 RN [1] sequence from N.A.
 RC STRAIN=cv. Cheyenne;
 RX MEDLINE=89098419; PubMed=2563152;
 RA Anderson O.D., Greene F.C., Yip R.E., Halford N.G., Shewry P.R., Malpica-Romero J.M.;
 RT "Nucleotide sequences of the two high-molecular-weight glutelin genes from the D-genome of a hexaploid bread wheat, *Triticum aestivum* L. cv Cheyenne";
 RL Nucleic Acids Res. 17:461-462 (1989).
 RN [2] revisions, sequence from N.A.
 RC STRAIN=cv. Cheyenne;
 RA Anderson O.D.;
 RL Submitted (PEB-1997) to the EMBL/GenBank/DDBJ databases.
 CC -1- FUNCTION: GLUTENINS ARE THE HIGH MOLECULAR WEIGHT SEED STORAGE PROTEIN OF WHEAT ENDOSPERM. THOUGHT TO BE RESPONSIBLE FOR THE VISCO-ELASTIC PROPERTY OF WHEAT DOUGH.
 CC -1- SUBUNIT: DISULFIDE-BRIDGE LINKED AGGREGATES.
 CC -1- MISCELLANEOUS: GLUTENINS ARE CODED BY SEVERAL GENES ON EACH OF THE GROUP 1 CHROMOSOMES OF WHEAT.
 CC -1- MISCELLANEOUS: THE MATURE PROTEIN IS CHARACTERIZED BY A LARGE NUMBER OF WELL PRESERVED REPEATS OF THE TWO MOTIFS: GQPGQ AND GOOGGOGOGGYPYS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.sib.ch/announce/> or send an email to license@ebi.sib.ch).
 CC EMBL; X12928; CAA31395.1; -.
 CC DR PIR; S02262; InterPro; IPR01419; Glutinin.
 CC DR PRINTS; PF03157; Glutinin_hmw_1.
 CC DR PRINTS; PRO00210; GLUTININ.
 CC KW Seed Storage Protein; Repeat; Multigene family; Signal.
 CC FT SIGNAL 1 21
 FT CHAIN 22 839 GLUTENIN, HIGH MOLECULAR WEIGHT SUBUNIT Dx5.
 FT DOMAIN 131 801 REPEATS.
 SQ SEQUENCE 839 AA; 89359 MW; 0F14E1106D552643 CRC64;

Query Match 15.3%; Score 99.5%; DB 1; Length 839;
 Best Local Similarity 23.5%; Pred. No. 3.5;
 Matches 31; Conservative 22; Mismatches 57; Indels 22; Gaps 5;

QY 4 RGRSTYRP-----RPRRYVPE-----MIGPMRPEQFS---DEVEPATPEEGE 44
 DB 491 QGQPGYPTSPQQSGQGQQPGQGYDPTSPQQPGQGQ 550

QY 45 PATQRDPAAQEGED-EGASIGQGPRAHSQEGHPQTGCECDGPDQEMDPENPEEV 103
 DB 551 QPGQLQQPAQDQQGQQLAQGQQQPAQVQQGQGQQQPAQGQQGQQQPAQGQQQ 608

QY 104 VKTPBERGEKQS 115
 DB 609 QQPAQGQQGQ 620

Qy	61	EGASAGGCPKPEAHSSQEAGHPQTGCECENGPDGQEMDPPNPEVKTPPESEKQSQC 116	1	MWSGRGRTYRPRRYVEPPEMIGPMRPEQFSDEVEPATQDPAAAEQED 6
Db	61	EGASAGGCPKPEAHSSQEAGHPQTGCECENGPDGQEMDPPNPEVKTPPESEKQSQC 116	1	MGFPLRLIYRRPMLIVEE-----QPDEVESPTQSDSTPAEERD 4
RESULT 2				
Q8WTP9		PRELIMINARY;	PRT;	111 AA.
ID	Q8WTP9;			
AC				
DT	01-MAR-2002	(TREMBLrel. 20; Last sequence update)		
DT	01-MAR-2002	(TREMBLrel. 20; Last annotation update)		
DT	01-MAR-2002	(TREMBLrel. 20; Last annotation update)		
GN	XAGE-3			
OS	Homo sapiens (Human).			
OC	Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN				
RP	SEQUENCE FROM N.A.			
RA	Zendman A.J.W., van Kraats A.A., Weidle U.H., Ruiter D.J., Van Muilen G.N.P.; "Expression profile and alignment of the XAGE family of cancer/testis associated genes"; Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.			
RT	"Expression profile of members of the XAGE cancer/testis family"; Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.			
RT	"Novel human cDNA clones with function of inhibiting cancer cell growth."; Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.			
RL	AJ318893; CAC83014.1; -.			
DR	AJ318881; CAC83008.1; -.			
SQ	SEQUENCE 111 AA; 12302 MW; P4A3261E6F7ACF5B CRC64;			
Query Match	42.0%; Score 271.5; DB 4; Length 111; Best Local Similarity 50.0%; Pred. No. 2.5e-16; Matches 59; Conservative 13; Mi smatches 35; Indels 11; Gaps 3;			
Qy	1 MSWRGRGRTYRPRRYVEPPEMIGPMRPEQFSDEVEPATQDPAAAEQ- 57			
Db	1 MIWRGRGRTYRPRRSVPPELGP-----LEPGDDEPQQBEPPTS RDPAQQR 52			
Qy	58 GEDEGASAGQGPKEAHSSQEAGHPQTGCECENGPDGQEMDPPNPEVKTPPESEKQSQ 115			
Db	53 EBDQGAATQVDPLEADIQELQSQTGGCNGPDDQGTILPKSEQFKMDPEGGDRQO 110			
RESULT 3				
Q9BSS7		PRELIMINARY;	PRT;	146 AA.
ID	Q9BSS7			
AC	Q9BSS7;			
DT	01-JUN-2001 (TREMBLrel. 17; Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17; Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19; Last sequence update)			
DB	G antigen, family B, 1 (prostate associated).			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	SEQUENCE FROM N.A.			
RC	TISSUE=SKIN;			
RA	Strausberg R.; Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.			
RL	BC004661; AAH04811.1; -.			
DR	SEQUENCE 146 AA; 16150 MW; E667BA94D89BDB2 CRC64;			
Qy	41.2%; Score 266; DB 4; Length 146; Best Local Similarity 40.0%; Pred. No. 9.9e-16; Matches 66; Conservative 7; Mi smatches 30; Indels 54; Gaps 3;			

RT	"Expression profile and alignment of the XAGE family of cancer/testis associated genes."	DB	XAGE-4 protein (Fragment).
RT	Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.	GN	XAGE-4.
RN	[3]	OS	Homo sapiens (Human).
RL	SEQUENCE FROM N.A.	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
RP	TISSUE=EWING SARCOMA;	OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RC	Zendman A.J.W.; van Kraats A.A., Weidle U.H., Ruiter D.J., Van Muijen G.N.P.	NCBI_TaxID	9606;
RA	"Expression profile of members of the XAGE cancer/testis antigen family."	RN	[1]
RT	Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.	SEQUENCE FROM N.A.	
RL	EMBL; BC00932; AAH09232.1; -.	RA	Zendman A.J.W.; van Kraats A.A., Weidle U.H., Ruiter D.J., Van Muijen G.N.P.
DR	EMBL; Au318891; CAC8125.1; -.	RT	"Expression profile and alignment of the XAGE family of cancer/testis associated genes."
DR	EMBL; Au318892; CAC8125.1; JOINED.	RT	Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AJ318890; CAC83092.1; -.	DR	EMBL; AJ318895; CAC83092.1; -.
DR	SEQUENCE 111 AA; 12354 MW; 38C9332C5BA0E14 CRC64;	FT	NON-TER 1
SQ	Query Match 35.5%; Score 229.5; DB 4; Length 111; Best Local Similarity 45.8%; Pred. No. 1e-12; Mismatches 15; Indels 38; Gaps 3; Matches 54; Conservative	FT	NON-TER 1
Qy	1 MSWGRSTRPRPRYVEPPMIGPMRPEQFSDEVPAPPAQEG 58	SQ	SEQUENCE 69 AA; 7621 MW; 69BF06B9B400E650 CRC64;
Db	1 MSWGRSTRPRPRPSSLQPLIGAM-----LEPSVPEQEPPTESQDPGKQR 52	Query Match 19.3%; Score 124.5; DB 4; Length 69; Best Local Similarity 33.7%; Pred. No. 0.00063; Mismatches 10; Indels 23; Gaps 3; Matches 30; Conservative	
Qy	59 E-DEGASAGQPKPAHSQEQGHQPTGCEDGPGDQEMDPNPNEETKTPBEGEKQSQ 115	Qy	19 PPEMIGPMRPEQFSDEVPAPQDPAAQEGIDEGASAGQGPRAEHSQ 76
Db	53 EDDQGAAEIQVDPLEADLQLCQTKTGDGGTQVKGKILPKAENHKMPERAGEGKSQ 110	Db	2 PPELIGPM-----LEPSDEPQQBEEPPTRSDPPTV-----POLETDLQ 40
RESULT 6			
QBWMMI	PRELIMINARY;	PRT;	108 AA.
ID	QBWMMI;	ID	O36421; PRELIMINARY;
AC	QBWMMI;	AC	QBWMMI; PRT; 1300 AA.
DT	01-MAR-2002 (TREMBLrel. 20, Created)	DT	036421; PRT; 1300 AA.
DT	01-MAR-2002 (TREMBLrel. 20, Last sequence update)	DT	01-JAN-1998 (TREMBLrel. 05, Created)
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)	DT	01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DE	XAGE-5 protein.	DE	Putative immediate early protein.
GN	XAGE-5	GN	Alcelaphine herpesvirus 1 (wildboar herpesvirus).
OS	Homo sapiens (Human).	OS	Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Gammaherpesvirinae; Rhadinovirus.
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	OC	NCBi_TaxID:35252;
NCBI_TaxID	9606;	RN	[1]
RN	SEQUENCE FROM N.A.	SEQUENCE FROM N.A.	
RP	Zendman A.J.W.; van Kraats A.A., Weidle U.H., Ruiter D.J., Van Muijen G.N.P.	RP	SEQUENCE FROM N.A.
RA	"Expression profile and alignment of the XAGE family of cancer/testis associated genes."	RC	STRAIN=C500;
RT	Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.	RX	PubMed=97404659;
RL	EMBL; Au318894; CAC83015.1; -.	RA	Ensler A., Pflanz R., Fleckenstein B.,
DR	SEQUENCE 108 AA; 12077 MW; E3CD91E5C9241628 CRC64;	RT	"Primary structure of the alcelaphine herpesvirus 1 genome.";
SQ	Query Match 30.8%; Score 199; DB 4; Length 108; Best Local Similarity 42.4%; Pred. No. 4.1e-10; Mismatches 14; Indels 4; Matches 50; Conservative	RL	J. Virol. 71:6517-6525(1997).
Qy	1 MSWGRSTRPRPRYVEPPMIGPMRPEQFSDEVPAPQDPAAQEG 58	DR	EMBL; AF005370; AAC58118.1;
Db	1 MSWGR---RYPRRCLRLAQLVGP-----LEPSVPEQEPPTESQDPGKQR 49	DR	InterPro; IPR000087; Collagen;
Qy	59 E-DEGASAGQPKPAHSQEQGHQPTGCEDGPGDQEMDPNPNEETKTPBEGEKQSQ 115	SQ	SEQUENCE 1300 AA; 128183 MW; 40F99FD244F34577 CRC64;
Db	50 EDDQGAAEIQVNLADLQELQSQTGDCGDSPDVYQKLPKSEQKMPPEGEGKQ 107	Qy	17 VEPPENIGPMRPEQFSDEVPAPQDPAAQEG-----PATQRQDPAAQEG-59
RESULT 7			
QBWMMO	PRELIMINARY;	PRT;	69 AA.
ID	QBWMMO;	ID	QBWMMO 114
AC	QBWMMO;	AC	P-BEGECOS
DT	01-MAR-2002 (TREMBLrel. 20, Created)	DT	646 P-BEGECOS 652
DT	01-MAR-2002 (TREMBLrel. 20, Last sequence update)	DT	RESULT 9 O18563

Wed Apr 9 18:05:33 2003

us-09-782-745-27.rspt

Page 6

Search completed: April 7, 2003, 18:15:45
Job time : 36 secs

PT detect the presence of PAGE-4 in cell samples or body tissues
 XX Disclosure: Figure 1a; 63pp; English.

CC PAGE-4 is a gene preferentially expressed in normal male and female reproductive tissues e.g. prostate, testis, fallopian tube, uterus and placenta, as well as in prostate cancer, testicular cancer and uterine cancer. This expression pattern makes it a target for diagnosis and for vaccine based therapies of such neoplasms.

CC An isolated PAGE-4 peptide which induces a cytotoxic T lymphocyte response when bound to a major histocompatibility complex (MHC) class I molecule or the isolated PAGE-4 protein can be used in immunogenic compositions to raise a cytotoxic T lymphocyte response against cells expressing PAGE-4 including cancer cells of the prostate, uterus and testis. The nucleic acids encoding PAGE-4 or PAGE-4 peptide fragments can also be used in these compositions.

CC Antibodies against PAGE-4 and its peptide fragments can be used in detecting the presence of PAGE-4 and PAGE-4 coding sequences in cell samples or body tissues. The presence of PAGE-4 in tissues which are not related to reproduction can be indicative of the spread of cancerous reproductive tissue. PAGE-4 can also be used to raise antibodies which are then used as the targeting group of immunoconjugates comprising toxins used in therapeutic applications. This has applications for drug delivery systems. The PAGE polypeptide shares sequence similarity with the GAGE and MAGE family of proteins.

XX Sequence 116 AA;

Query	Match	Score	Length
Qy	1 MSWGRSTYPRPRVYVEPPMIGPMRPEQFDEVEPATPEGEPATQRDPAAAQEGED 60	100 0%	DB 21;
Db	1 MSWGRSTYPRPRVYVEPPMIGPMRPEQFDEVEPATPEGEPATQRDPAAAQEGED 60	100 0%	Pred. No. 1e-52;
Qy	61 EGASAGQGPKEAHSQEQHQTGCBEDGPQEMDPNPEEVKTPEEGEKQSQC 116	0	Mismatches 0;
Db	61 EGASAGQGPKEAHSQEQHQTGCBEDGPQEMDPNPEEVKTPEEGEKQSQC 116	0	Gaps 0;

XX Example 13; Fig 5; 60pp; English.
 XX The present sequence represents a GAGE-2 tumour rejection antigen precursor (TRAP). The protein is expressed in a number of tumours. In contrast the only normal tissue which expresses GAGE TRAP protein is testis. Several GAGE TRAPS have been identified (see AAV1871-21).
 CC The major difference between these proteins and GAGE-1 is the absence of a stretch of 143 bases located at position 379 to 521 of the GAGE-1 TRAP sequence. The rest of the sequences show mismatches at various positions, with the exception of GAGE-3 whose 5' end is totally different from the other GAGE cDNAs for the first 112 bases. This region of GAGE-3 cDNA contains a long repeat and a hairpin structure.
 CC The antigens can be used to diagnose melanomas, characterised by expression of a TRAP or presentation of a tumour rejection antigen.
 CC Antigens shed into blood or urine can be observed and then used to confirm a diagnosis of melanoma using cytoytic T cell clone proliferation methodologies. Other uses for the processed peptides, include HLA-typing assays for, e.g. skin graft or organ transplants.

XX Sequence 116 AA;

Query	Match	Score	Length
Qy	1 MSWGRSTYPRPRVYVEPPMIGPMRPEQFDEVEPATPEGEPATQRDPAAAQEGED 60	98 8%;	DB 19;
Db	1 MSWGRSTYPRPRVYVEPPMIGPMRPEQFDEVEPATPEGEPATQRDPAAAQEGED 60	99 1%;	Pred. No. 5.5e-52;
Qy	61 EGASAGQGPKEAHSQEQHQTGCBEDGPQEMDPNPEEVKTPEEGEKQSQC 116	0	Mismatches 0;
Db	61 EGASAGQGPKEAHSQEQHQTGCBEDGPQEMDPNPEEVKTPEEGEKQSQC 116	0	Gaps 0;

XX Sequence 116 AA;

XX Best Local Similarity 99.1%; Indels 0; Gaps 0;
 CC Matches 115; Conservative 0; Mismatches 0;
 CC Query Match 3 AAY83162 standard; Protein: 117 AA.
 CC ID AAY83162
 CC AC AAY83162;
 CC DT 24-JUL-2000 (first entry)
 CC DE GAGE4 polypeptide.
 CC KW PAGE-4; MAGE; GAGE; reproduction; testis; prostate; fallopian tube; uterus; placenta; cancer; major histocompatibility complex; MHC; CRL; cytotoxic T lymphocyte; immune response; antibody; drug delivery; immunoconjugate.
 CC KW Homo sapiens.
 CC OS Homo sapiens.
 CC PN WO200012706-A1.
 CC PD 09-MAR-2000.
 CC XX 31-AUG-1999; 99WO-US20046.
 CC XX PR 01-SEP-1998; 98US-0098993.
 CC XX PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 CC XX PI Pastan I, Brinkmann U, Vassatzis G, Lee B;
 CC XX PA (LUDWIG) LUDWIG INST CANCER RES.

RESULT 2
 ID AAW47599
 XX AAW47599 standard; Protein: 116 AA.

XX AC AAW47599;
 XX DT 30-JUL-1998 (first entry)

XX DE GAGE-2 tumour rejection antigen precursor.

XX KW GAGE tumour tumour rejection antigen precursor; TRAP; tumour; diagnosis; melanoma; antigen; cytolytic T cell clone proliferation; HLA-typing assay.

XX OS Homo sapiens.

XX FH Location/Qualifiers
 FT Misc-difference 38
 /note= "Ala encoded by GAG"
 FT Misc-difference 39
 /note= "Thr encoded by CCT"

XX FT WO9749417-A1.
 XX PD 31-DEC-1997.
 XX PT 23-JUN-1997; 97WO-US10850.
 XX PR 24-JUN-1996; 96US-0669161.
 XX PA

DR WPI; 2000-237869/20.
 XX PT Isolated PAGE-4 protein and peptide used for inducing a cytotoxic T lymphocyte response and for raising antibodies which can be used to detect the presence of PAGE-4 in cell samples or body tissues
 XX PS Disclosure; Figure 1a; 63pp; English.

XX PAGE-4 is a gene preferentially expressed in normal male and female reproductive tissues e.g. prostate, testis, fallopian tube, uterus and placenta, as well as in prostate cancer, testicular cancer and uterine cancer. This expression pattern makes it a target for diagnosis and for vaccine based therapy of such neoplasms. An isolated PAGE-4 peptide which induces a cytotoxic T lymphocyte response when bound to a major histocompatibility complex (MHC) class I molecule or the isolated PAGE-4 protein can be used in immunogenic compositions to raise a cytotoxic T lymphocyte response against cancer cells of the prostate, uterus and testis. The nucleic acids encoding PAGE-4 or PAGE-4 peptide fragments can also be used in these compositions. Antibodies against PAGE-4 and its peptide fragments can be used in detecting the presence of PAGE-4 and PAGE-4 coding sequences in cell samples or body tissues. The presence of PAGE-4 in tissues which are not related to reproduction can be indicative of the spread of cancerous reproductive tissue. PAGE-4 can also be used to raise antibodies which are then used as the targeting group of immunoconjugates comprising toxins used in therapeutic applications. This has applications for drug delivery systems. The PAGE polypeptide shares sequence similarity with the GAGE and MAGE family of proteins.

XX Sequence 117 AA;

Query Match 96.5%; Score 623.5; DB 21; Length 117;
 Best Local Similarity 97.4%; Pred. No. 1.2e-50;
 Matches 114; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

XX Query 1 MSWRGRST-YRPRPRRYVPPERMIGPMRPEQSDEVEPATPPEGEPATQRDPAAAQEGE 59
 Db 1 MSWRGRST-YRPRPRRYVPPERMIGPMRPEQSDEVEPATPPEGEPATQRDPAAAQEGE 60

XX Query 60 DEGASAGQGPKPKEAHSQEGHPOQTGCCECDGPDQEMDPNPNEEVYKTPPEGEKEKOSQC 116
 Db 61 DEGASAGQGPKPKEAHSQEGHPOQTGCCECDGPDQEMDPNPNEEVYKTPPEGEKEKOSQC 117

XX Sequence 117 AA;

Query Match 96.5%; Score 623.5; DB 21; Length 117;
 Best Local Similarity 97.4%; Pred. No. 1.2e-50;
 Matches 114; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

XX Query 1 MSWRGRST-YRPRPRRYVPPERMIGPMRPEQSDEVEPATPPEGEPATQRDPAAAQEGE 59
 Db 1 MSWRGRST-YRPRPRRYVPPERMIGPMRPEQSDEVEPATPPEGEPATQRDPAAAQEGE 60

XX Query 60 DEGASAGQGPKPKEAHSQEGHPOQTGCCECDGPDQEMDPNPNEEVYKTPPEGEKEKOSQC 116
 Db 61 DEGASAGQGPKPKEAHSQEGHPOQTGCCECDGPDQEMDPNPNEEVYKTPPEGEKEKOSQC 117

RESULT 4

AY83163
 XX ID AAY83163 standard; Protein; 117 AA.
 XX AC AAY83163;

XX DT 24-JUL-2000 (First entry)

XX DE GAGES polypeptide.

XX KW PAGE-4; MAGE; GAGE; reproduction; testis; prostate; fallopian tube; uterus; placenta; cancer; major histocompatibility complex; MHC; CTL; cytotoxic T lymphocyte; immune response; antibody; drug delivery; immunooncjugate.

XX OS Homo sapiens.

XX PN WO200012706-A1.

XX PD 09-MAR-2000.

XX PF 31-AUG-1999; 99WO-US20046.

XX PR 01-SEP-1998; 98US-009993.

XX PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PI Pastan I, Brinkmann U, Vassatzis G, Lee B;
 XX DR WPI; 2000-237869/20.
 XX PT Isolated PAGE-4 protein and peptide used for inducing a cytotoxic T lymphocyte response and for raising antibodies which can be used to detect the presence of PAGE-4 in cell samples or body tissues
 XX PS Disclosure; Figure 1a; 63pp; English.
 XX PAGE-4 is a gene preferentially expressed in normal male and female reproductive tissues e.g. prostate, testis, fallopian tube, uterus and placenta, as well as in prostate cancer, testicular cancer and uterine cancer. This expression pattern makes it a target for diagnosis and for vaccine based therapy of such neoplasms. An isolated PAGE-4 peptide which induces a cytotoxic T lymphocyte response when bound to a major histocompatibility complex (MHC) class I molecule or the isolated PAGE-4 protein can be used in immunogenic compositions to raise a cytotoxic T lymphocyte response against cancer cells of the prostate, uterus and testis. The nucleic acids encoding PAGE-4 or PAGE-4 peptide fragments can also be used in these compositions. Antibodies against PAGE-4 and its peptide fragments can be used in detecting the presence of PAGE-4 and PAGE-4 coding sequences in cell samples or body tissues. The presence of PAGE-4 in tissues which are not related to reproduction can be indicative of the spread of cancerous reproductive tissue. PAGE-4 can also be used to raise antibodies which are then used as the targeting group of immunoconjugates comprising toxins used in therapeutic applications. This has applications for drug delivery systems. The PAGE polypeptide shares sequence similarity with the GAGE and MAGE family of proteins.
 XX SQ Sequence 117 AA;
 XX Query Match 95.9%; Score 619.5; DB 21;
 Best Local Similarity 96.6%; Pred. No. 2.9e-50;
 Matches 113; Conservative 2; Mismatches 1; Indels 1; Gaps 1;
 XX Qy 1 MSWRGRST-YRPRPRRYVPPERMIGPMRPEQSDEVEPATPPEGEPATQRDPAAAQEGE 59
 Db 1 MSWRGRST-YRPRPRRYVPPERMIGPMRPEQSDEVEPATPPEGEPATQRDPAAAQEGE 60
 XX Qy 60 DEGASAGQGPKPKEAHSQEGHPOQTGCCECDGPDQEMDPNPNEEVYKTPPEGEKEKOSQC 116
 Db 61 DEGASAGQGPKPKEAHSQEGHPOQTGCCECDGPDQEMDPNPNEEVYKTPPEGEKEKOSQC 117
 XX SQ Sequence 117 AA;
 XX Query Match 95.9%; Score 619.5; DB 21;
 Best Local Similarity 96.6%; Pred. No. 2.9e-50;
 Matches 113; Conservative 2; Mismatches 1; Indels 1; Gaps 1;
 XX Qy 1 MSWRGRST-YRPRPRRYVPPERMIGPMRPEQSDEVEPATPPEGEPATQRDPAAAQEGE 59
 Db 1 MSWRGRST-YRPRPRRYVPPERMIGPMRPEQSDEVEPATPPEGEPATQRDPAAAQEGE 60
 XX Qy 60 DEGASAGQGPKPKEAHSQEGHPOQTGCCECDGPDQEMDPNPNEEVYKTPPEGEKEKOSQC 116
 Db 61 DEGASAGQGPKPKEAHSQEGHPOQTGCCECDGPDQEMDPNPNEEVYKTPPEGEKEKOSQC 117
 XX RESULT 5
 XX ID AAW47601 standard; Protein; 117 AA.
 XX AC AAW47601;
 XX DT 30-JUL-1998 (First entry)
 XX DE GAGE-4 tumour rejection antigen precursor.
 XX KW GAGE tumour tumour rejection antigen precursor; TRAP; tumour; diagnosis; melanoma; antigen; cytolytic T cell clone proliferation; HLA-typing assay.
 XX OS Homo sapiens.
 XX FH Key Location/Qualifiers
 XX FT Misc-difference 39
 XX FT /note= "Ala encoded by GAG"
 XX FT Misc-difference 40
 XX FT /note= "Thr encoded by CCT"
 XX PN WO979417-A1.
 XX PD 31-DEC-1997.

XX	23-JUN-1997;	97WO-US10850.	PR	01-SEP-1998;	98US-0098993.
XX	24-JUN-1997;	96US-0669161.	PA	(USSH) US DEPT HEALTH & HUMAN SERVICES.	
PR			XX		
XX			PI	Pastan I, Brinkmann U, Vassatzis G, Lee B;	
PA	LUDWIG INST CANCER RES.		XX		
XX			DR	2000-237869/20.	
PI	Boon-Falleur T, Debacker O, Van Den Eynde B;		XX		
XX	DR	Isolated PAGE-4 protein and peptide used for inducing a cytotoxic T lymphocyte response and for raising antibodies which can be used to detect the presence of PAGE-4 in cell samples or body tissues	PT		
XX	N-PSDB; AAV18719.		PT		
PT	Isolated nucleic acid encoding GAGE tumour rejection antigen precursor - processed by HLA-Cw6 molecules into peptides, useful to diagnose melanomas		PT		
PT	Example 13: Fig 5; 60pp; English.		PS		
XX	The present sequence represents a GAGE-4 tumour tumour rejection antigen precursor ('TRAP'). The protein is expressed in a number of tumours. In contrast the only normal tissue which expresses GAGE TRAP protein is testis. Several GAGE TRAPs have been identified (see AAV1871-7-21). The major difference between these proteins and GAGE-1 is the absence of a stretch of 143 bases located at position 379 to 521 of the GAGE-1 TRAP sequence. The rest of the sequences show mismatches at various positions, with the exception of GAGE-3 whose 5' end is totally different from the other GAGE cDNAs for the first 112 bases. This region of GAGE-3 cDNA contains a long repeat and a hairpin structure. The antigens can be used to diagnose melanomas, characterised by expression of a TRAP or presentation of a tumour rejection antigen. Antigens shed into blood or urine can be observed and then used to confirm a diagnosis of melanoma using cytolytic T cell clone proliferation methodologies. Other uses for the processed Peptides, include HLA-typing assays for, e.g. skin graft or organ transplants.		SQ	Sequence 117 AA;	
XX	Sequence 117 AA;		SQ	117 AA;	
QY	MSWRGRSTY-PRPRRYVRPPMPEIGPMRPEFQSDEVEPATPEEGEPATQRDPAAAQEGE 59		QY	MSWRGRSTY-PRPRRYVRPPMPEIGPMRPEFQSDEVEPATPEEGEPATQRDPAAAQEGE 59	Score 615.5; DB 21; Length 117; Best Local Similarity 95.3%; Pred. No. 6.9e-50; Matches 112; Conservative 3; Mismatches 1; Indels 1; Gaps 1;
Db	1 MSWRGRSTYWRPRRTVQQPENITGPMPEFQSDEVEPATPEEGEPATQRDPAAAQEGE 60		Db	1 MSWRGRSTYWRPRRTVQQPENITGPMPEFQSDEVEPATPEEGEPATQRDPAAAQEGE 60	
QY	60 DEGASAGQPKPEAHSQPOGHPTQCCCECDGPQEMDDPNPEEVTKTPEEGEKOSQC 116.		QY	60 DEGASAGQPKPEAHSQPOGHPTQCCCECDGPQEMDDPNPEEVTKTPEEGEKOSQC 116	
Db	61 DEGASAGQPKPEADSQBGHPQTKTPEEVTKTPEEGEKOSQC 117		Db	61 DEGASAGQPKPEADSQBGHPQTKTPEEVTKTPEEGEKOSQC 117	
XX	RESULT 6				
ID	AY83164		ID	AY83160	RESULT 7
AC	AY83164 standard; Protein; 117 AA.		ID	AY83160 standard; Protein; 118 AA.	
XX			XX		
AC	AY83164;		AC	AY83160;	
XX			XX		
DT	24-JUL-2000 (first entry)		DT	24-JUL-2000 (first entry)	
XX			XX		
DE	GAGE6 polypeptide.		DE	GAGE1 polypeptide.	
XX			XX		
KW	PAGE-4; MAGE; GAGE; reproduction; testis; prostate; fallopian tube; uterus; placenta; cancer; major histocompatibility complex; MHC; CTL; cytotoxic T lymphocyte; immune response; antibody; drug delivery; Homo sapiens.		KW	PAGE-4; MAGE; GAGE; reproduction; testis; prostate; fallopian tube; uterus; placenta; cancer; major histocompatibility complex; MHC; CTL; cytotoxic T lymphocyte; immune response; antibody; drug delivery; Homo sapiens.	
XX			KW		
OS	W0200012706-A1.		OS		
XX			XX		
PN	09-MAR-2000.		PN	WO200012706-A1.	
XX			PD	09-MAR-2000.	
PF	1-Aug-1999; 99WO-US20046.		XX		

PF	31-AUG-1999;	99WO-US20046.	
XX	01-SEP-1998;	98US-009893.	
PA	(USSH) US DEPT HEALTH & HUMAN SERVICES.		
PI	Pastan I, Brinkmann U, Vasmatzis G, Lee B;		
XX	WPI; 2000-237869/20.		
XX	Isolated PAGE-4 protein and peptide used for inducing a cytotoxic T lymphocyte response and for raising antibodies which can be used to detect the presence of PAGE-4 in cell samples or body tissues		
XX	Disclosure: Figure 1a; 63bp; English.		
CC	PAGE-4 is a gene preferentially expressed in normal male and female reproductive tissues e.g. prostate, testis, fallopian tube, uterus and placenta, as well as in prostate cancer, testicular cancer and uterine cancer. This expression pattern makes it a target for diagnosis and for vaccination based therapy of such neoplasms.		
CC	An isolated PAGE-4 peptide which induces a cytotoxic T lymphocyte response when bound to a major histocompatibility complex (MHC) class I molecule or the isolated PAGE-4 protein can be used in immunogenic compositions to raise a cytotoxic T lymphocyte response against cells expressing PAGE-4 including cancer cells of the prostate, uterus and testis. The nucleic acids encoding PAGE-4 or PAGE-4 peptide fragments can also be used in these compositions. Antibodies against PAGE-4 and its peptide fragments can be used in detecting the presence of PAGE-4 and PAGE-4 coding sequences in cell samples or body tissues. The presence of PAGE-4 in tissues which are not related to reproduction can be indicative of the spread of cancerous reproductive tissue. PAGE-4 can also be used to raise antibodies which are then used as the targeting group of immunoconjugates comprising toxins used in therapeutic applications. This has applications for drug delivery systems. The PAGE polypeptide shares sequence similarity with the GAGE and MAGE family of proteins		
XX	Sequence 118 AA;		
Query Match	94.9%;	Score 613; DB 21; Length 118;	
Best Local Similarity	97.4%;	Prd. No. 1.2e-49;	
Matches	111;	Conservative 1; Mismatches 2;	
Indels	0;	Gaps 0;	
Qy	1 MSWRGRSTYRPRRYVTPPEPPMIGPMRPEQFDEVPATPEEGEPATQRQDPAAAQGED 60		
Db	1 MSWRGRSTYRPRRYVTPPEPPMIGPMRPEQFSDVEPATPEEGEPATQRQDPAAAQGED 60		
Qy	61 EGASAGGQPKPEAHSSQFGHPQTGCGCBECDGPQEMPPNPNEVKTPEEGEKS 114		
Db	61 EGASAGGQPKPEADSQFGHPQTGCGCBECDGPQEMPPNPNEVKTPEEGEGRNS 114		
RESULT 8	A9W47602		
ID	A9W47602 Standard; Protein; 117 AA.		
XX			
AC	A9W47602;		
DT	30-JUL-1998 (first entry)		
DE	GAGE-5 tumour rejection antigen precursor.		
XX			
KW	GAGE tumour rejection antigen precursor; TRAP; tumour; diagnosis; melanoma; antigen; cytolytic T cell clone proliferation; HLA-typing assay.		
XX			
OS	Homo sapiens.		
FH	Location/Qualifiers		
FT	/note= "Ala encoded by GAG"		

FT	Misc-difference 40	/note= "Thr encoded by CCT"
XX	WO9749417-A1.	
XX	31-DEC-1997.	
PD	23-JUN-1997;	97WO-US10850.
XX	24-JUN-1996;	96US-0669161.
XX	(LUDW-) LUDWIG INST CANCER RES.	
PA	Boon-Falleur T, Debacker O, Van Den Eynde B;	
PI	WPI; 1998-076905/07.	
DR	N-PSDB; AAV18720.	
PT	Isolated nucleic acid encoding GAGE tumour rejection antigen precursor - processed by HLA-Cw6 molecules into peptides, useful to diagnose melanomas	
XX	Example 13: Fig 5; 60pp; English.	
PS	The present sequence represents a GAGE-5 tumour tumour rejection antigen precursor (TRAP). The protein is expressed in a number of tumours. In contrast the only normal tissue which expresses GAGE TRAP protein is testis. Several GAGE TRAPs have been identified (see AAV18717-21). The major difference between these proteins and GAGE-1 is the absence of a stretch of 143 bases located at position 379 to 521 of the GAGE-1 TRAP sequence. The rest of the sequences show mismatches at various positions, with the exception of GAGE-3 whose 5' end is totally different from the other GAGE cDNAs for the first 112 bases. This region of GAGE-3 cDNA contains a long repeat and a hairpin structure. The antigens can be used to diagnose melanomas, characterised by expression of a TRAP or presentation of a tumour rejection antigen. Antigens shed into blood or urine can be observed and then used to confirm a diagnosis of melanoma using cytolytic T cell clone proliferation methodologies. Other uses for the processed peptides, include HLA-typing assays for, e.g. skin graft or organ transplants.	
SQ	Sequence 117 AA:	
Query Match	94.8%; Score 612.5; DB 19;	Length 117;
Best Local Similarity	95.7%; Pred. No. 1.3e-49;	
Matches	112;	Conservative 2; Mismatches 2;
Indels	1;	Gaps 1;
Oy	1 MSWRGRSTYRPRRYVTPPEPPMIGPMRPEQFSDVEPATPEEGEPATQRQDPAAAQGED 59	
Db	1 MSWRGRSTYRPRRYVTPPEPPMIGPMRPEQFSDVEPATPEEGEPATQRQDPAAAQGED 60	
Qy	60 DEGASAGGQPKPEADSQFGHPQTGCGCBECDGPQEMPPNPNEVKTPEEGEKSQC 114	
Db	61 DEGASAGGQPKPEADSQFGHPQTGCGCBECDGPQEMPPNPNEVKTPEEGEGRNS 117	
RESULT 9	A9W47603	
ID	A9W47603 Standard; Protein; 117 AA.	
XX		
AC	A9W47603;	
DT	30-JUL-1998 (first entry)	
DE	GAGE-6 tumour rejection antigen precursor.	
XX		
KW	GAGE tumour rejection antigen precursor; TRAP; tumour; diagnosis; melanoma; antigen; cytolytic T cell clone proliferation; HLA-typing assay.	
XX		
OS	Homo sapiens.	
FH	Location/Qualifiers	
FT	/note= "Ala encoded by GAG"	
	Location/Qualifiers	

	Location/Qualifiers
XX	
FH	
Misc-difference 39	
PT	/note= "encoded by GAG"
FT	
Misc-difference 40	
FT	/note= "encoded by CCT"
XX	
PN	WO9749417-A1.
XX	
PD	31-DEC-1997.
XX	
PF	23-JUN-1997;
XX	97WO-US10850.
PR	24-JUN-1996;
XX	96US-0669161.
PA	(LUDWIG) LUDWIG INST CANCER RES.
XX	
PI	Boon-Falleur T, Debacker O, Van Den Eynde B;
XX	
DR	WPI: 1998-076905/07.
DR	N-PSDBB; AAV05540.
XX	
PT	Isolated nucleic acid encoding GAGE tumour rejection antigen precursor - processed by HLA-Cw6 molecules into peptides, useful to diagnose melanoma
PT	
PT	Example 7: Fig 5: 60pp; English.
XX	
PS	The present sequence represents a GAGE-1 tumour tumour rejection antigen precursor (TRAP). The protein is expressed in a number of tumours. In contrast the only normal tissue which expresses GAGE TRAP protein is testis. Several GAGE TRAPs have been identified (see AAV18717-21).
CC	The major difference between these proteins and GAGE-1 is the absence of a stretch of 143 bases located at position 379 to 521 of the GAGE-1 TRAP sequence. The rest of the sequences show mismatches at various positions, with the exception of GAGE-3 whose 5' end is totally different from the other GAGE cDNAs for the first 112 bases. This region of GAGE-3 cDNA contains a long repeat and a hairpin structure.
CC	The antigens can be used to diagnose melanomas, characterised by expression of a TRAP or presentation of a tumour rejection antigen.
CC	Antigens shed into blood or urine can be observed and then used to confirm a diagnosis of melanoma using cytolytic T cell clone proliferation methodologies. Other uses for the processed peptides, include HLA-typing assays for, e.g. skin graft or organ transplants.
XX	
SQ	Sequence 138 AA:
	Query Match 92.6%; Score 598; DB 19; Length 138;
	Best Local Similarity 99.1%; Pred. No. 3.5e-48;
	Matches 108; Conservative 0; Mismatches 1; Indels 0; Gaps 0
Qy	1 MSWRGRSTYRPRRYVEPPERMIGPMRPECFSDSDEVEPATPEBGEPATQRQDPAAAQEQQED 60
Db	1 MSWRGRSTYRPRRYVEPPERMIGPMRPECFSDSDEVEPATPEBGEPATQRQDPAAAQEQQED 60
Qy	61 EGASAGQGPKPRAHSOEOGHQPTGCECDGPQMDPNNPVEEVKTPPEE 109
Db	61 EGASAGQGPKPPEADSQBOGHQPTGCECDGPQMDPNNPBEVKTPEE 109
RESULT 11	
AAU84812	
ID	AAU84812 standard; Protein; 138 AA.
XX	
AC	AAU84812;
XX	
DT	08-MAY-2002 (First entry)
XX	
DE	Human GAGE-1 consensus sequence.
XX	
KW	Savine; vaccine; cancer; viral infection; HIV; hepatitis C virus;
KW	viral infection; human immunodeficiency virus; melanoma; bacterial infection; Salmonella; Legionella; parasitic infection;
KW	

KW PAGE-4; MAGE; GAGE; reproduction; testis; prostate; fallopian tube;
 KW uterus; placenta; cancer; major histocompatibility complex; MHC; CTL;
 KW cytotoxic T lymphocyte; immune response; antibody; drug delivery;
 KW immunoconjugate.
 XX
 OS Homo sapiens.
 XX WO2001019197-A1.
 PN
 XX PD 29-NOV-2001.
 XX
 XX 25-MAY-2001; 2001WO-AU00622.
 XX
 PR 26-MAY-2000; 2000AU-0007761.
 PA (AUSU) UNIV AUSTRALIAN NAT.
 XX
 PI Thomson SA, Ramshaw IA,
 XX
 DR WPI; 2002-147575/19.
 XX
 PT New synthetic polypeptides having several different segments of at least one parent polypeptide linked together differently compared to the linkage in the parent polypeptide, for inducing immune response against a pathogen or cancer.
 PT
 PS Example 3; Fig 27; 364pp; English.
 XX
 CC The invention relates to a new synthetic polypeptide (I) comprising several different segments of at least one parent polypeptide linked together in a different relationship relative to their linkage in the parent polypeptide to impede, abrogate or otherwise alter at least one function associated with the parent polypeptide and for inducing an immune response against a pathogen or cancer. Also included are a synthetic polynucleotide encoding and a computer system for designing the synthetic polypeptides. The synthetic polypeptides and polynucleotides are referred to as a Savine. The synthetic polypeptide is useful for modulating immune responses preferably directed against a pathogen or a cancer, (e.g., cancers of the lung, breast, ovary, cervix, colon, head and neck, pancreas, prostate, stomach, bladder, kidney, bone liver, oesophagus, brain, testicle, uterus), as potentiating agents. Compositions comprising the polypeptide may be used in the treatment or prophylaxis against viral (such as infections caused by HIV (human immunodeficiency virus), hepatitis, influenza, Japanese encephalitis virus, Epstein-Barr virus and respiratory syncytial virus), bacterial (e.g., infections caused by Neisseria, Meningococcal, Haemophilus, Salmonella, Streptococcal, Legionella and Mycobacterium or parasitic (e.g., infections caused by Plasmodium, Schistosoma, Leishmania, Trypanosoma, Toxoplasma and Giardia) infections. The present sequence is a consensus sequence for a parent protein used to design a savine of the invention.
 XX Sequence 138 AA;
 SQ Score 92.6%; DB 23; Length 138;
 Best Local Similarity 99.1%; Pred. No. 3.5e-48;
 Matches 108; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Query Match Score 59.5%; DB 21; Length 118;
 Best Local Similarity 92.3%; Pred. No. 4.1e-48;
 Matches 109; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

QY 1 MSWGRSTYRPRRYVEPEPMIGMRPEQSDEVPATBEGEPATQRDPAAQEGED 60
 Db 1 MSWGRSTYRPRRYVEPEPMIGMRPEQSDEVPATBEGEPATQRDPAAQEGED 60

QY 61 EGASAGQGPKPPEAHSQEQGHPQTGCEDPGQEMDPNPPEEVKTPBE 109
 Db 61 EGASAGQGPKPPEAHSQEQGHPQTGCEDPGQEMDPNPPEEVKTPBE 109

RESULT 12
 AAW83161 ID AAW83161 standard; Protein; 118 AA.
 XX AC AAW83161;
 XX DT 24-JUL-2000 (first entry)
 DE GAGE3 polypeptide.

RESULT 13
 AAW47600 ID AAW47600 standard; Protein; 118 AA.
 XX AC AAW47600;
 DT 30-JUL-1998 (first entry)

XX DE GAGE-3 tumour rejection antigen precursor.

XX KW GAGE tumour rejection antigen precursor; TRAP; tumour; diagnosis; melanoma; antigen; cytolytic T cell clone proliferation; HLA-typing assay.

XX OS Homo sapiens.

XX FH KeyY Location/Qualifiers

FT Misc-difference 40 /note= "Ala encoded by GAG"

FT Misc-difference 41 /note= "Thr encoded by CCT"

FT WO9749417-A1.

XX PN PD 31-DEC-1997.

XX PR 23-JUN-1997; 97WO-US10850.

XX PR 24-JUN-1996; 96US-0669161.

XX PA (LUDWIG) LUDWIG INST CANCER RES.

XX PI Boon-Falleur T, Debacker O, Van Den Eynde B;

XX DR WPI; 1998-076905/07.

DR N-PSDB; AAV18718.

XX PT Isolated nucleic acid encoding GAGE tumour rejection antigen precursor - processed by HLA-Cw6 molecules into peptides, useful to diagnose melanomas

XX PS Example 13; Fig 5; 60PP; English.

CC The present sequence represents a GAGE-3 tumour tumour rejection antigen precursor (TRAP). The protein is expressed in a number of tumours. In contrast the only normal tissue which expresses GAGE TRAP protein is testis. Several GAGE TRAPS have been identified (see AAV1871-21).

CC The major difference between these proteins and GAGE-1 is the absence of a stretch of 143 bases located at position 379 to 521 of the GAGE-1 TRAP sequence. The rest of the sequences show mismatches at various positions, with the exception of GAGE-3 whose 5' end is totally different from the other GAGE cDNAs for the first 112 bases. This region of GAGE-3 cDNA contains a long repeat and a hairpin structure.

CC The antigens can be used to diagnose melanomas, characterised by expression of a TRAP or presentation of a tumour rejection antigen.

CC Antigens shed into blood or urine can be observed and then used to confirm a diagnosis of melanoma using cytolytic T cell clone proliferation methodologies. Other uses for the processed peptides, include HLA-typing assays for, e.g. skin graft or organ transplants.

XX SQ Sequence 118 AA:

Query Match 91.3%; Score 589.5; DB 19; Length 118;

Best Local Similarity 94.7%; Pred. No. 1.8e-47;

Matches 108; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

Qy 4 RGRSTYRVEPPMIGPMRPEQSDEVEPATPPEGEPATQRQDPAAQEGEDEG 62

Db 5 RGKSTYYWPRLPRRYVQPPEVIGPMRPEQQSDEVEPATPPEGEPATQRQDPAAQEGEDEG 64

Qy 63 ASAGQGPKEAHSQEQQHPTQGEEPEVKTPEEGKQSQC 116

Db 65 ASAGQGPKEAHSQEQQHPTQGEEPEVKTPEEGKQSQC 118

RESULT 14
AAG02123 ID Human nootropic; immunosuppressant; cyrostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

XX XX XX AC

XX DT 06-OCT-2000 (first entry)

XX DE Human secreted protein, SEQ ID NO: 6204.

XX KW Human; 5'-EST; expressed sequence tag; secreted protein; cDNA isolation; gene therapy; chromosome mapping.

XX OS Homo sapiens.

XX PN EP1033401-A2.

XX PD 06-SEP-2000.

XX PF 21-FEB-2000; 2000EP-02000610.

XX PR 26-FEB-1999; 99US-01224487.

(GEST) GENSET.

XX PA

XX PI Dumas Milne Edwards J, Ducleart A, Giordano J;

XX DR WPI; 2000-500381/45.

XX DR N-PSDB; AAC02129.

XX PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX PS Claim 13; SEQ ID 6204: 71PP + CD-ROM; English.

XX The present sequence is a polypeptide encoded by one of a large number of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs derived from 30 different tissues. EST sequences usually correspond mainly to the 3' untranslated region (UTR) of the mRNA because they are often obtained from oligo-dT primed cDNA libraries. Such ESTs are not well suited for isolating cDNA sequences derived from the 5' ends of mRNAs and even in those cases where longer cDNA sequences have been obtained, the full 5' UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and chromosome mapping procedures. They are used to obtain upstream regulatory sequences and to design expression and secretion vectors.

XX SQ Sequence 76 AA:

Query Match 61.9%; Score 400; DB 21;

Best Local Similarity 97.3%; Pred. No. 4.9e-30;

Matches 73; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 22 MIGMRPEQSDEVEPATPPEGEPATQRQDPAAQEGEDEG 81

Db 1 MIGMRPEQSDEVEPATPPEGEPATQRQDPAAQEGEDEG 60

RESULT 15
AAM39588 ID Human polypeptide SEQ ID NO 2733.

XX KW Human; nootropic; immunosuppressant; cyrostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

XX DT 22-OCT-2001 (first entry)

XX DE

KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.

XX Homo sapiens.

WO200153312-A1.

XX PD 26-JUL-2001.

XX PF 26-DEC-2000; 2000WO-US34263.

XX PR 21-JAN-2000; 2000US-0488725.

XX PR 25-APR-2000; 2000US-0552317.

XX PR 09-JUL-2000; 2000US-0598042.

XX PR 19-JUL-2000; 2000US-0620312.

XX PR 03-ANG-2000; 2000US-0653450.

XX PR 14-SEP-2000; 2000US-0662191.

XX PR 19-OCT-2000; 2000US-0693016.

XX PR 29-NOV-2000; 2000US-0727344.

XX PA (HYSEQ INC.

XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

PI Wang J, Wang Z, Wehrman T, Xu C, Xue AU, Yang Y, Zhang J;

PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX DR WPI; 2001-442253/47.

DR N-PSDB; AAI58744.

XX PT Novel nucleic acids and polypeptides, useful for treating disorders such as central nervous system injuries -

XX PS Example 4; SEQ ID NO 2733; 10078pp; English.

XX CC The invention relates to human nucleic acids (AAI57798-AAI61360) and the encoded polypeptides (AAM38442-AAM42213) with nocrotropic, immunosuppressant and cytostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: Immune system suppression, Activ/inhibit activity, chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, leukaemias and C.N.S disorders.

CC Note: The sequence data for this patent did not form part of the printed specification.

XX Sequence 111 AA:

Query Match Score 271.5; DB 22; Length 111;
 Best Local Similarity 50.0%; Prod. No. 6.5e-18; Mismatches 13; Indels 11; Gaps 3;

Matches 59; Conservative 13; Mismatches 35; Indels 11; Gaps 3;

QY 1 MSWRGISTYPRPRRYVEPPENITGPMREBEQSDEVEPA-TPEBEGEPATORQDPAAAAB- 57

Db 1 MIWRGISTYPRPRRSVPPBLGPM-----LEPDDEBEPQEPPTESRDAPGER 52

QY 58 GEDEGASAGOGKPEAHNSQEOGHPTGGCFCEDGPDGQEMDPNPPEEVTKPEEGEKOSQ 115

Db 53 BEDQGAETQVBDLEADDQELSOKTGGCNGPDDQQRILPKSEQFMPEGGBDRQQ 110

